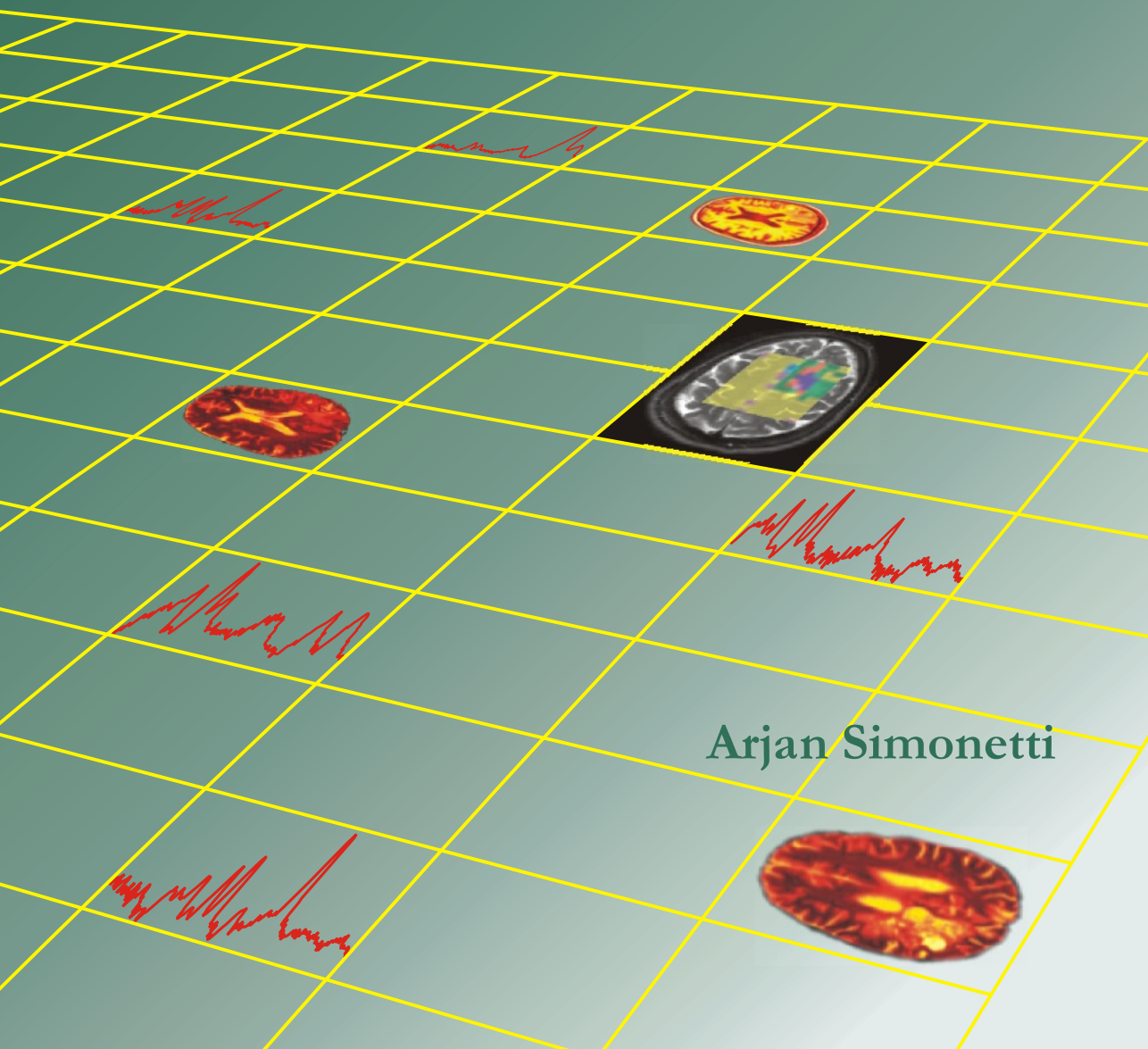
Investigation of brain tumor classification and its reliability using chemometrics on MR spectroscopy and MR imaging data



Arjan Simonetti

Investigation of brain tumor classification and its reliability using chemometrics on MR spectroscopy and MR imaging data

Investigation of brain tumor classification and its reliability using chemometrics on MR spectroscopy and MR imaging data

een wetenschappelijke proeve op het gebied van de
Natuurwetenschappen, Wiskunde en Informatica

Proefschrift

ter verkrijging van de graad doctor aan de Katholieke Universiteit Nijmegen, op gezag van de Rector Magnificus Prof Dr. C.W.P.M. Blom, volgens besluit van het College van Decanen in het openbaar te verdedigen op donderdag 15 april 2004 des namiddags om 1.30 uur precies

door
Arjan Willem Simonetti
geboren op 8 januari 1972
te Hengelo

Promotores: Prof. Dr. L.M.C. Buydens
Prof. Dr. A. Heerschap

Manuscriptcommissie: Prof. Dr. C. Segebarth
Université Joseph Fourier, Frankrijk

Prof. Dr. P. van Espen
Universitaire Instellingen Antwerpen, België

Dr. M. van der Graaf

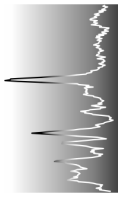
Cover design by Arjan W. Simonetti

The studies presented in this thesis were performed at the Laboratory for Analytical Chemistry, University of Nijmegen, The Netherlands. Financial support was obtained from the EU funded INTERPRET project (IST-1999-10310).

ISBN 90-9017814-7

Printed by Ponsen & Looijen B.V., Wageningen

Contents



Chapter 1

Introduction

Arjan W. Simonetti

1.1 The diagnosis of brain tumors

The occurrence of brain tumors is one of the leading causes of cancer death in young adults and the second fastest growing cause of cancer death among people over age 65 [1]. Persons with a brain tumor often have symptoms like gradual loss of movement or sensation in an arm or leg, loss of vision or speech difficulty. In hospital, a radiological investigation may demonstrate an abnormality in the brain. Since the symptoms of the patient can easily be confused with other disorders of the brain, e.g. infarct or abces, the first step is to identify the disorder as being a tumor. If a brain tumor has been diagnosed, the next very important step is to identify the type of tumor and possibly the tumor grade. The tumor type is dependent on the type of cell the tumor originates from. In this thesis the focus is on three types of tumors: oligodendroglioma, astrocytoma and meningioma. The first two tumor types arise from the brains supportive tissue, and are collectively called gliomas. Meningiomas are tumors that arise from the leptomeninges (tissue that serves as the lining of the brain). The tumor grade indicates the level of tumor malignancy. Tumors are graded on their growth rate, vascularity (blood supply), presence of a necrotic center, invasive potential (border distinctness) and similarity to normal cells. Malignant tumors may contain several grades of cells. The most malignant grade of cell found determines the grade for the entire tumor, even if most of the tumor is a lower grade.

It is rarely possible to decide upon the tumor type and grade with absolute certainty from the clinical history and radiological investigations. There are a number of specialized tests to diagnose brain tumors. A neurological examination checks: eye movement, sensory-motoric and muscular responses, reflexes, hearing, tactile and sensory sensation, movement, balance and co-ordination. Next, special imaging techniques and laboratory tests can be used, that detect the presence of a tumor and provide clues about its location. The most commonly used techniques for diagnosing brain tumors are[1]:

a) **Computed Tomography (CT)**, which uses X-ray radiation to create a detailed image of the body's tissues and structure;
b) **Magnetic Resonance Imaging (MRI)**, which uses a magnetic field rather than X-rays to create an image, and can often distinguish more accurately between healthy and abnormal tissue;

c) **Positron Emission Tomography (PET)**, which provides a picture of brain activity rather than structure by measuring local levels of injected compounds, such as glucose or methionine, that have been labeled with a radioactive tracer;

d) **Biopsy** and characterization by a pathologist is a final step in confirming the diagnosis of a brain tumor. A biopsy is a surgical procedure in which a small sample of tissue is taken from the suspected tumor. Then, a more precise diagnosis, revealing the type and grade of the tumor can be obtained by histopathology.

In spite of the quality of the aforementioned noninvasive radiological tools, often a biopsy is still required before a treatment plan can be developed. Brain biopsy is an invasive surgical procedure and has a significant risk of mortality. Biopsy does not always provide the correct answer, since it is difficult to collect the tissue sample from the exact location. Often, the tissue sample is heterogeneous, which complicates the identification of the tumor type by the pathologist. Furthermore, a biopsy provides only local information from the tumor, while tumors are known to be heterogeneous and infiltrative. Therefore it is very important to develop new, noninvasive, radiological tools that can aid in the diagnosis of tumor type and grade.

1.2 Can MR spectroscopy assist in tumor diagnosis?

A new technique, that is not commonly used yet in the clinic is **Magnetic Resonance Spectroscopy (MRS)**. MRS provides completely different information as compared to MRI, but can be obtained with most 1.5 Tesla MRI instruments used for routine radiology. During MRS, spectra are obtained that contain information about the concentrations and chemical properties of an extended set of metabolites. Two techniques are commonly applied for this purpose[2]. First, the MR spectra can be obtained from a specific selected region (volume element or voxel) in the brain. This is called single voxel spectroscopy (SVS). In brain tumor diagnosis, the voxel normally includes the tumorous area. Next to SVS, also multiple voxels can be measured simultaneously. Normally, this implies the acquisition of spectra from a two- (or three-) dimensional grid. Each voxel contains an MR spectrum, that provides metabolic information about the volume it is measured from. The method is

normally referred to as magnetic resonance spectroscopic imaging (MRSI) or chemical shift imaging (figure 1.1). MRSI can be especially helpful in the identification of heterogeneity in the tumorous part of the brain, since changes in the tissue structure can be assessed on a metabolic level.

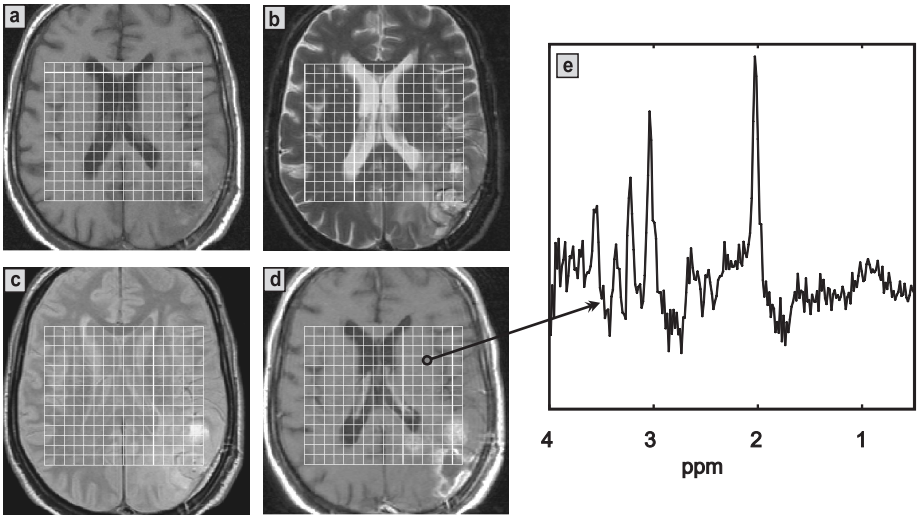
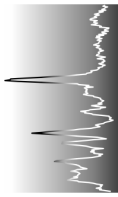


Figure 1.1. Example of the information obtained by MRI and MRSI. *MR images with different contrasts are commonly obtained for diagnosing brain tumors. Examples of a T_1-weighted, T_2-weighted, proton density weighted and T_1-weighted MR image after administration of a Gadolinium-DTPA tracer are given in figure a to d. A grid is displayed over the images. From each location in the grid, one complete proton MR spectrum is obtained during MRSI. An example of such a proton MR spectrum is given in figure e.*

MRS can be performed with several MR sensitive nuclei. Next to the proton nucleus (^{1}H), other potentially useful nuclei for biomedical applications include ^{31}P, ^{19}F and ^{13}C. The use of ^{31}P-MRS has been found important in studies of energy metabolism, as the content of ATP, phosphocreatine and inorganic phosphate during exercise can be followed[3]. It can also be used to calculate intracellular pH[4], which is an important parameter in the study of tumors. Also the signals of phosphomonoesters in tumors can be measured by ^{31}P-MRS. These signals are found to be raised in tumors compared with normal tissue.



The metabolic information obtained by ^{1}H-MRS has proven to be important in the investigation of brain tumors[5, 6]. Several metabolite concentrations have been found to be lowered or increased in tumors compared with normal tissue. This change in concentration can be used for specific tumor determination. The proton nucleus has a high MR sensitivity and its natural abundance is high. A large number of (brain) metabolites contain protons, that generate signals at specific frequencies in the MR signal. The number of signals is so high that these actually crowd the MR spectrum and the overlap of signals is manifest.

To influence the presence of signals from metabolites in the proton MR spectrum, acquisition protocols can be used that use different acquisition parameters. One of these parameters, the echo-time, has a direct influence on the signal of metabolites in the MR spectrum. If the echo-time is long, the number of signals from metabolites is lower and the baseline of the spectrum is often flat. Proton MR spectra acquired with a short echo-time contain more signals but can have a distorted baseline due to metabolic components that resonate in a large range of the spectrum. Short echo-time spectra contain more information, but the extraction of this information might be difficult compared with long echo-time spectra.

In principle, the data from proton MR spectroscopy can assist in the noninvasive clinical diagnosis of brain tumors directly. The spectral pattern of spectra obtained from tumors if different from spectra obtained form healthy brain tissue. Also the patterns of spectra obtained from different tumor types and tumor grades are dissimilar. A well trained clinician could use the information of the patterns to substantiate the diagnosis. However, the appearance of spectroscopic data is very different compared to imaging data. Clinicians are used to diagnose patients on basis of imagers, not spectra. Furthermore, the amount of data generated, especially during MRSI, hampers the diagnosis by clinicians. Going through all the data would simply be too time consuming. Therefore, there is a need for the development of statistical models that condense the data and that provide an output that is easily interpreted by the radiologists and other clinicians. Optimally, the output should be some kind of a brain image in order to map the most malignant tumor part as well as the tumor boundaries and to compare the results with the conventionally obtained MR images. The main goal of the development of statistical models would be to aid the clinician in the diagnosis of brain tumors, by improving the accessibility of the information within the MR

spectra. Ultimately, the statistical interpretation of noninvasively obtained data from multiple methodologies could lead to a situation in which the surgical procedure, the biopsy, is not necessary anymore. This would be very beneficial to the patient, clinician and insurance, since a biopsy is tremendously stressful for the patient, time-consuming, expensive, and to some extent error-prone regarding the final clinical diagnosis.

1.3 The information content of MRI and MRS

MRI has become one of the most important noninvasive aids in clinical diagnosis of brain tumors, because it enables the radiologist to assess the anatomical location, morphology and size of the tumor tissue with a relatively high spatial resolution. Stacked images are recorded throughout the total volume of the brain in order to localize the tumor optimally. Also MR images with different contrasts are recorded, as the acquisition of only one type of image may not show the desired information. The contrasts are normally obtained by the acquisition of T_1-weighted, T_2-weighted and proton density weighted images. This is often followed by a T_1-weighted image after intravenous administration of Gadolinium-DTPA, to assess the blood brain barrier viability (figure 1.1). Next to this, also more modern MRI methods may be used, like diffusion-weighted imaging and functional MRI. During diffusion-weighted imaging, the water movement within the brain tissue is analyzed at the molecular level. An abnormality can be detected and visualized on diffusion-weighted images, since the normal water movement is restricted. With functional MRI it is possible to visualize which specific brain structure participates in a specific function, and is therefore important in the assessment of the function of the brain tissue surrounding a tumor.

The proton MR spectra obtained by MRS contain information from important brain metabolites like glutamate, myo-inositol, choline, creatine, N-acetyl aspartate (NAA), lactate and fatty acids[7]. NAA is considered to be a neuronal marker of which the levels are decreased in brain tumors. The choline resonance arises from a group of compounds that are involved in membrane synthesis and degradation. Increased choline is often a marker of tumor aggressiveness. The creatine signal reflects compounds involved in energy metabolism and may be decreased in brain tumors. Lactate is normally below detection limit, but it is often increased in brain tumors. The function of myo-inositol is largely unknown, but its signal appears to be altered in brain

tumors. The signals from fatty acids are known to be increased in several brain tumors, especially high grades, and are an indicator for necrosis (cell death). Glutamate is an important neurotransmitter in brain and may be elevated in oligodendrogliomas[8] and other tumors.

The classical way to use the metabolic information stored in the spectral pattern for the diagnosis of brain tumors is to quantitate the intensity of several signals in the MR spectra. The most simple methodologies integrate spectral regions of interest, based on the assumption that the underlying points are highly (lorentzian or gaussian) correlated and contain information from one single metabolite. More specialized routines fit the spectrum in the spectral domain using a basis set of experimental model spectra[9]. Also fitting in the time domain, using a priori knowledge is used[10]. A significant statistical change in one of the quantitated metabolite levels is often used as a marker for malignancy of the tumor. However, in most cases merely univariate methods are used, while the availability of information is highly suited for a multivariate approach.

1.4 Extracting information from spectroscopic data using chemometrics

Another, fundamentally different way of collecting information from spectra is to use chemometrics[11]. Chemometrics is a discipline, that has the goal to obtain information and knowledge from data in the field of chemistry. It explores chemical data with the use of mathematics, statistics and formal logic[12]. Chemometric models are used to mine information from multivariate imaging[13, 14], spectroscopic data[15, 16], and are applicable to a broad gamut of chemical and physical problems. These models can be applied to data obtained from MRI and MRS also, since these methods acquire data using the same basic methodology as methods in for example high-resolution NMR in biochemistry.

Several investigators have used chemometric approaches on MRI[17] data or MRSI data to diagnose or, less difficult, localize brain tumors[5, 18-26]. Howells[27, 28] has been one of the first who used chemometric methods on MR data to discriminate between tumors. She prepared chemical extracts from tumor tissue, which were analyzed with high resolution NMR. Several chemometric methods were used to analyze the data. Preul[5] has introduced chemometrics in the classification of brain tumors using noninvasive methods. He used MRSI

data in combination with a linear discriminant analysis technique to discriminate between the five most common types of brain tumors in adults. MRSI was used to measure the spatial distribution of the (relative) concentrations of several metabolic compounds, resulting in so-called metabolic maps or images (figure 1.2). Then, the multivariate data were used to classify the tumors. Metabolic maps have also been used to assess the heterogeneity of tumors[29, 30], since it is important to know the heterogeneity in the tumor and surrounding tissue for clinical diagnosis and evaluation of disease.

In order to increase the accuracy of classification, several classification approaches have been developed. A hierarchical tree-like decision system[31] uses multiple binary classifiers, each optimized for a specific decision (e.g. healthy vs. abnormal, or malignant vs. benign). Another approach is to use consensus diagnosis[20], in which several independent classifiers work in parallel. If the majority of the classifiers yield the same classification result, then this consensus is accepted as being the 'true' class with a high accuracy.

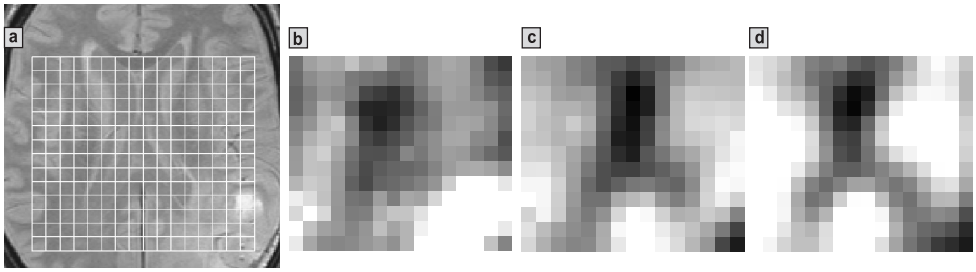


Figure 1.2. Metabolic maps constructed from MRSI data. *The MR image shows a tumor in the lower right of the brain (a). To investigate the metabolic changes in the brain, metabolic maps of choline (b), creatine (c) and N-acetyl aspartate (d) are constructed from the MRSI data. Bright colors represent a high intensity, and thus a high concentration of metabolic compound.*

1.5 The challenges for chemometrics to improve tumor classification

Four stages can be defined in the strategy of tumor classification using chemometrics in combination with MRI and MRS data. Each will be described shortly, followed by the challenges for chemometrics.

The first stage comprises of a selection of training samples for specific tumor classes to be classified. Preferably, these samples are selected from a database of samples that are acquired in the same unique way. The training set has to contain sufficient samples for each tumor class. Ideally, for each class also an independent test set should be available, in order to test the accuracy of the classifier. To obtain enough samples from a large number of tumors, databases have to be developed[32] that not only contain MR imaging and spectroscopic data, but also clinical information. All data should be easily accessible to the researcher, and should be validated. Although this seems to have nothing in common with chemometrics, it does! The selection of data by the researcher is important for the final classification result. Therefore, software has to be developed that guides the researcher through the large amount of data in the databases. Methods have to be developed that deal with missing or corrupted data, and which perform automated quality control. Only then it can be thoroughly investigated whether for example the contribution of image and clinical data to spectroscopic data has a positive effect on tumor classification.

The second step consists of all kinds of preprocessing on the data[33]. This may involve several steps. The most common classic methods in the processing of ^{1}H-MRS data are; a) Eddy current correction[34], b) residual water signal removal (water filtering)[35], c) Fourier transformation, d) first order phase correction, and e) normalization.

Preprocessing of the MRS data is a very important issue, especially since signals with a low signal to noise ratio are important in tumor diagnosis. The challenges for chemometrics in this area are the development of fully automated procedures that exhibit robustness and that can identify outliers. This is important, because the user in the clinic will probably not have expertise in data processing. It is also important in a statistical sense, since several pre-processing steps have to be executed in a cascade and error propagation has to be kept to a minimum. Furthermore, methods have to be developed that prepare data from several sources, e.g. MRS, MRI, CT, PET

and clinical information. This will facilitate the combination of data from different radiological methods into one input data set. One of the most difficult problems in this field that needs investigation, deals with the best matching of the differences in spatial resolution between MRI and MRSI data. Also the matching between images from different modalities is a difficult problem, since pixels from images with the same pixel location should contain information from the same spatial location.

The third step in the strategy of tumor classification involves feature reduction of the highly dimensional spectroscopic data. During feature reduction, variables are selected from the data that are meaningful for the classification. This is especially important if spectroscopic data is used in the classification of brain tumors, since in this case the number of variables often exceeds the number of objects. Feature reduction should be performed in such a way that the information that optimally discriminates the tumor classes is kept. Unfortunately, this is not always a straightforward task. As mentioned, one can use classic quantitative procedures based on a priori information or use chemometrics to perform feature reduction. There is a major challenge foreseen for chemometrics in this part of tumor classification, since a lot of methods are available that have proven to be very successful. One can think of methods like principal components analysis, factor analysis and genetic algorithms[28, 36-38]. Each of these methods has its advantages and disadvantages. Probably, for each classification problem the best feature reduction algorithm needs to be found separately.

The last step involves the actual classifier. If the feature reduction has been sufficient and retained much of the class discriminating information, then the direct distance to the class centers (centroids) or the k-nearest neighbors method can be used. The data can also be used as input for methods that further reduce the dimensionality by finding optimal separable directions (e.g. linear discriminant analysis or quadratic discriminant analysis). Furthermore, density functions, neural networks, genetic algorithms, factor analysis and support vector machines can be used to construct the classifier. The number of methods is steadily growing[19, 20, 23, 28, 39, 40], and selection of the best one is becoming a major task. If data from imaging modalities is used (this also includes MRSI), the newly developed classification method also needs to deal with the spatial distribution. Chemometrics has been used to develop classifiers which take the classification result of neighboring voxels into

account. The tumor is normally located in one specific region of the brain, so this could improve classification.

Since the clinician is interested as much in the reliability of classification as in the classification itself, the methods developed should also be able to generate probabilities of classification. With the use of chemometrics it is possible to calculate these probabilities. It should be investigated whether it is possible to generate robust reliability estimates, parallel to the classification. If this is indeed the case, then chemometrics can play an important part in the acceptance of automated brain tumor classification in the clinic.

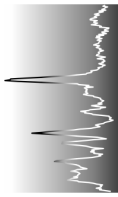
1.6 Outline of the thesis

In this thesis, chemometric methods are investigated that are used on proton MRSI data and MRI data, in order to develop a reliable and noninvasive brain tumor diagnosis model. As discussed in Chapter 1.5, there are challenges for chemometrics in all stages of the strategy of automated brain tumor classification using MRSI and MRI data. To find out what the most important challenges to solve were, the methods normally used during each stage of brain tumor classification were investigated. Brain tumor classification is a very difficult issue. A good classification result can only be obtained if each stage in the strategy is thoroughly investigated and dealt with. As a result, new methods were developed for the preprocessing, feature reduction and classification procedures used in brain tumor classification. With respect to the classification, a model was developed that also provides probabilities of classification. Probabilities of classification give insight into the reliability of the classification. This is important, since the clinicians who will use the system have to be convinced that the classification results are meaningful. Hopefully, the developed methods described in this thesis will be of help in better understanding tumor morphology, type and grade and may be of help in finding the best treatment and follow-up of patients with a brain tumor.

In order to test all investigated chemometric approaches, a semi-automated classification system has been developed that uses information obtained from MRSI data as well as MRI data. At the beginning of this study no automated preprocessing procedures were available. A lot of time has been invested in the development of such procedures to preprocess the data in the best possible way. This has been followed by the development of a segmentation

and classification strategy. It was investigated whether the segmentation had additional value in obtaining a more reliable classification result. The classification strategy was developed such that it calculated a probability of classification for each classification result. These probabilities were presented image like, to facilitate an easy interpretation in the clinic. Both, the results of classification and the results of the probability maps have been evaluated. Finally, several feature selection approaches have been tested in their ability to condense the data, so as to have the best input for the classification strategy.

MRI and MRSI basics are presented in Chapter 2, focusing on the role they play in the diagnosis of brain tumors. Some processing techniques for MR images and MR spectroscopic images are also addressed in this section. In Chapter 3, a new preprocessing technique is described based on a standard preprocessing method (eddy current correction[34]), to improve the automated preprocessing of MRSI spectra. The application of a semi-automatic unsupervised clustering technique (mclust[41]) to *in vivo* MRI and MRSI data of patients with a brain tumor is investigated and described in Chapter 4. This clustering technique is used to segment the brain. The segments are classified using a new classification algorithm. This algorithm, described in Chapter 5, also calculates the probability of classification for each segment. Besides, the classification and probability of classification of each individual voxel are calculated. The two classification images, one from the classified segments and one from the classified individual voxels, provide a consensus diagnosis of the brain tumor, which is described in Chapter 6. In Chapter 7 several feature reduction methods are compared. It is examined which one provides the best input to classify brain tumors of different type and grade, normal tissue and cerebro-spinal fluid. Next to the comparison of these feature reduction methods also the influence of MRI information and the water signal intensity is considered. Finally, concluding remarks and recommendations are given for future research and researchers.



References

[1] http://www.tbts.org/welcome.htm

[2] Sauter R, Schneider M, Wicklow K, Kolem H. Current status of clinically relevant techniques in magnetic resonance spectroscopy. Electromedia 1992; **60**: 32-54.

[3] Gadian DG. *NMR and its applications to living systems.* second edition, Oxford University Press, 1995

[4] Negendank W, Studies of human tumors by MRS: a review. NMR in Biomedicine 1992; **5**: 302-324.

[5] Preul MC, Caramanos Z, Collins DL, Villemure J, Leblanc R, Olivier A, Pokrupa R, Arnold DL. Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy. Nature Medicine 1996; **2**: 323-325.

[6] Nelson SJ, Vigneron DB, Dillon WP. Serial evaluation of patients with brain tumors using volume MRI and 3D ^{1}H MRSI. NMR in Biomedicine 1999; **12**: 123-138.

[7] Govindaraju V, Young K, Maudsley AA. Proton NMR chemical shifts and coupling constants for brain metabolites. NMR in Biomedicine 2000; **13**: 129-153.

[8] Rijpkema M, Schuuring J, van der Meulen Y, et al. Characterization of oligodendrogliomas using short echo time ^{1}H MR spectroscopic Imaging. NMR in biomedicine 2003; **16(1)**: 12-18.

[9] Provencher SW. Estimation of Metabolite Concentrations from Localized in vivo Proton NMR spectra. Magnetic Resonance in Medicine 1993; **30**: 672-679.

[10] de Beer R, van Ormondt D. Analysis of NMR data using time domain fitting procedures. NMR Basic Princ. Progr. 1992; **26**: 201-248.

[11] Beebe KR, Seasholtz MB, *Chemometrics A Practical Guide.* John Wiley and Sons, New York, 1998.

[12] Vandeginste BGM, Massart DL, Buydens LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J. *Handbook of Chemometrics and Qualimetrics: Part B.* Elsevier, 1998; Chapter 33.

[13] Esbensen KH, Geladi P. Strategies of multivariate image analysis (MIA). Chemom. Intell. Lab. Syst. 1989; **7**:67-86

[14] Geladi P, Grahn H. *Multivariate image analysis*. John Wiley and Sons, Chichester, 1996.

[15] Salzer R, Steiner G, Mantsch HH, Mansfield J, Lewis EN. Infrared and Raman imaging of biological and biomimetic samples. Fresenius' journal of analytical chemistry 2000; **366**:712-726.

[16] Schaeberle MD, Morris HR, Turner II JF, Treado PJ. Raman chemical imaging spectroscopy. Analytical Chemistry 1999; **71**:175A-181A.

[17] Clarke LP, Velthuizen RP, Camacho MA, Heine JJ, Vaidyanathan M, Hall LO, Thatcher RW, Silbiger ML. Mri Segmentation: methods and application. Magnetic Resonance Imaging 1995; **13**:343-368.

[18] Preul MC, Caramanos Z, Leblanc R, Villemure JG, Arnold DL. Using pattern analysis of in vivo proton MRSI data to improve the diagnosis and surgical management of patients with brain tumors. NMR in Biomedicine 1998; **11**: 192-200.

[19] Hagberg G, Burlina AP, Mader I, Roser W, Radue EW, Seelig J. In vivo proton MR spectroscopy of human gliomas: definition of metabolic coordinates for multi-dimensional classification. Magnetic Resonance in Medicine 1995; **34**: 242-252.

[20] Somorjai RL, Nikulin AE, Pizzi N, Jackson D, Scarth G, Dolenko B, Gordon H, Russell P, Lean CL, Delbridge L, Mountford CE, Smith ICP. Computerized consensus diagnosis: A classification strategy for the robust analysis of MR spectra. I. Application to 1H spectra of thyroid neoplasms. Magnetic Resonance in Medicine 1995; **33**: 257-263.

[21] Tate AR, Griffiths JR, Martinez-Pérez I, Moreno A, Barba I, Cabañas ME, Watson D, Alonso J, Bartumeus F, Isamat F, Ferrer I, Vila F, Ferrer E, Capdevila A, Arús C. Towards a method for automated classification of 1H MRS spectra from brain tumours. NMR in Biomedicine 1998; **11**: 177-191.

[22] Tate AR, Majós C, Moreno A, Howe FA, Griffiths JR, Arús C. Automated classification of short echo time in in vivo 1H brain tumor spectra: a multicenter study. Magnetic Resonance in Medicine 2003; **49**: 29-36.

[23] El-Deredy W. Pattern recognition approaches in biomedical and clinical magnetic resonance spectroscopy: a review. NMR in Biomedicine 1997; **10**: 99-124.

[24] Meyerand ME, Pipas JM, Mamourian A, Tosteson TD, Dunn JF. Classification of biopsy-confirmed brain tumors using single-voxel MR spectroscopy. American Journal Neuroradiology 1999; **20**: 117-123.

[25] Szabo de Edelenyi F, Rubin C, Estève F, Grand S, Décorps M, Lefournier V, Le Bas J, Rémy C. A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors: nosologic images. Nature Medicine 2000; **6**: 1287-1289.

[26] Herminghaus S, Dierks T, Pilatus U, Möller-Hartmann W, Wittsack J, Marquardt G, Labisch C, Lanfermann H., Schlote W, Zanella FE. Determination of histopathological tumor grade in neuroepithelial brain tumors by using spectral pattern analysis of in vivo spectroscopic data. Journal of Neurosurgery 2003; **98**: 74-81.

[27] Howells SL, Maxwell RJ, Griffiths JR. Classification of tumour 1H NMR spectra by pattern recognition. NMR in biomedicine 1992; **5**: 59-64.

[28] Howells SL, Maxwell RJ, Peet AC, Griffiths JR. An investigation of tumor 1H nuclear magnetic resonance spectra by the application of chemometric techniques. Magnetic Resonance in Medicine 1992; **28**: 214-236.

[29] Furuya S, Naruse S, Ide M, Morishita H, Kizu O, Ueda S, Maeda T. Evaluation of metabolic heterogeneity in brain tumors using 1H-chemical shift imaging method. NMR in Biomedicine 1997; **10**: 25-30.

[30] Segebarth C, Balériaux D, Luyten PR, den Hollander JA. Detection of metabolic heterogeneity of human intracranial tumors in vivo by 1H NMR spectroscopic imaging. Magnetic Resonance in Medicine 1990; **13**: 62-76.

[31] Minguillon J, Tate AR, Arús C, Griffiths JR. Classifier combination for in vivo magnetic resonance spectra of brain tumours. Lecture Notes in Computer Science 2002; **2364**: 282-292.

[32] http://carbon.uab.es/INTERPRET/interpret_home.shtml

[33] in 't Zandt HJA, van der Graaf M, Heerschap A. Common processing of in vivo MR spectra. NMR in biomedicine 2001; **14**: 224-232.

[34] Klose U. In vivo proton spectroscopy in presence of eddy currents. Magnetic Resonance in Medicine 1990; **14**: 26-30.

[35] Vanhamme L, Fierro RD, Van Huffel S, de Beer R. Fast removal of residual water in proton spectra. Journal of Magnetic Resonance 1998; **132**: 197-203.

[36] Holmes E, Foxall PJD, Nicholson JK, et al. Automatic data reduction and pattern recognition methods for analysis of 1H nuclear magnetic resonance spectra of human urine from normal and pathological states. Analytical Biochemistry 1994; **220**: 284-296.

[37] Windig W, Antalek A. Resolving nuclear magnetic resonance data of complex mixtures by three-way methods: Examples of chemical solutions and the human brain. Chemometrics and intelligent laboratory systems 1999; **46**: 207-219.

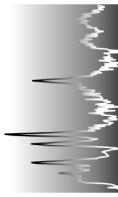
[38] Candolfi A, De Maesschalck R, Jouan-Rimbaud D, Hailey PA, Massart DL. The influence of data pre-processing in the pattern recognition of excipients near infrared spectra. Journal of pharmaceutical biomedical analysis 1999; **21**: 115-132.

[39] Hagberg G. From magnetic resonance spectroscopy to classification of tumors. A review of pattern recognition methods. NMR in biomedicine 1998; **11**: 148-156.

[40] Somorjai RL, Dolenko B, Nukulin A, et al. Distinguishing normal from rejecting renal allografts: application of a three-stage classification strategy to MR and IR spectra of urine. Vibrational Spectroscopy 2002; **28**: 97-102.

[41] McLachlan GJ, Peel D. *Finite mixture models*. John Wiley & Sons, New York, 2000.

Chapter 2



Better brain imaging with chemometrics‡

Han Witjes, Arjan W. Simonetti, and Lutgarde Buydens

Laboratory for Analytical Chemistry, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

‡ Published in Analytical Chemistry 2001; **73**(19), p 548A – 556A

Introduction

One of the many tangential benefits of the computer revolution has been the development of chemometrics, which has been adopted for many application in analytical chemistry. Chemometrics uses statistical and mathematical methods to unravel chemical problems. As described in various textbooks, chemometrics plays a decisive role in process analytical chemistry, both in monitoring and control application, and is increasingly used in areas such as analyzing images and optimizing molecular structures on the basis of their properties and activities[1-4].

Nuclear magnetic resonance (NMR) of molecules in living tissue is a new challenge for chemometrics. NMR can be performed as either spectroscopy (MRS) or imaging (MRI). MRI is an established medical technique for examining different internal organs like the brain, prostate, and heart and is an accepted clinical tool for imaging blood vessels and musculoskeletal tissue.

In particular, MRI is the most useful imaging modality for examining brain tumors. Acquiring MR images of representative cross-sectional slices of the head to detect and localize a brain tumor is a daily occurrence. One or more small tissue samples called biopsies are then removed from the localized tumor to determine identity and grade. Depending on the diagnosis, a patient next receives chemotherapy, radiation treatment, or undergoes surgery.

However, a biopsy is far from optimal for the diagnosis of brain tumors because it is invasive, subject to sampling errors, and provides no information about the tissue heterogeneity of the tumor. Therefore, MRI, MRS, and a combination of these techniques called magnetic resonance spectroscopic imaging (MRSI) are being developed in clinical trials in the hope of providing noninvasive brain tumor diagnosis.

Here lies the major challenge for chemometrics: No breakthroughs are expected in clinical MR examinations of brain tumor patients, because analysis and interpretation of the acquired MR data are still the major bottleneck. But the combination of MRI, MRSI, and chemometrics may advance the fast, noninvasive diagnosis of brain tumors.

In this feature we will explain some basis elements of MRI and MRSI in human brain studies and focus on the possibilities and added value that chemometrics can offer the techniques.

MRI basics

In figure 2.1a the spatial distribution of the proton NMR signal of water in a cross-sectional tissue slice of the human brain is shown as an image. The voxel (volume element) of each pixel in a MR image is ~0.01 cm^3. The strength of MRI is the possibility it offers the user to manipulate the image contrast by varying acquisition parameters. The so-called spin-echo imaging technique is used routinely in the clinical environment. It consists of (as any other MRI sequence) radio frequency pulses and magnetic field gradients, that are applied to a subject placed inside a homogeneous static magnetic field. The pulse sequence generates the proton NMR signal, while the gradient sequence spatially encodes the proton NMR signal.

The proton-density (ρ), and the spin-lattice (T_1) and spin-spin (T_2) relaxation times are tissue parameters that determine the NMR intensities of the different tissue types. The echo time (TE) and the repetition time (TR) are acquisition parameters that determine the ρ-, T_1-, and T_2-sensitivity of the acquired spin echo images. According to equation 1, the signal intensity (S) in each voxel can be described to a first-order approximation (i.e., assuming that TR >> TE) as:

$$S = \rho \bullet \left[1 - \exp^{-\frac{TR}{T_1}}\right] \bullet \exp^{-\frac{TE}{T_2}} \qquad 1$$

S is heavily ρ-weighted if TR is long with respect to T_1 and TE is short with respect to T_2. S is T_2-weighted if TR is long and TE is long, and S is T_1-weighted if TR is short and TE is long. Thus tissues which differ in ρ, T_1, or T_2 show different intensities in the ρ-weighted, T_1-weighted, or T_2-weighted images.

In the hospital, the ρ-weighted, T_1-weighted, and T_2-weighted images are commonly acquired with the spin echo technique with different combinations of TE and TR. These images are usually combined for the segmentation of normal brain tissue, i.e., the discrimination between white matter, gray matter, and cerebro-spinal fluid. However, additional feature images are required for the segmentation of brain tumors. Usually, a T_1-weighted spin-echo image after contrast agent administration is added to the standard images. The contrast agent shortens the T_1 of brain tissue that has a disrupted vascular structure as a result of tumor proliferation.

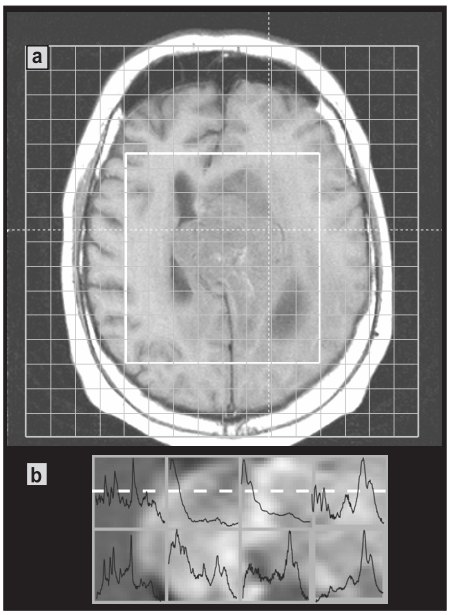
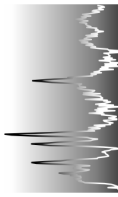


Figure 2.1. MR data acquired from a cross-sectional image slice through the brain tumor. *(a) The MR image shows the proton NMR signal of water and a high spatial resolution. The grid indicates the lower spatial resolution of the spectroscopic image. (b) Part of the spectroscopic image showing the MR spectra from adjacent spectroscopic voxels in the tumor region. Each spectrum shows the proton NMR signal of various metabolites like NAA, Cho, and Cr.*

Understanding MRSI of brain tumors

Figure 2.1b shows data from a MR spectroscopic image that shows the spatial distribution of the proton NMR signal of various metabolites in a cross-sectional tissue slice of a human brain. The voxel size in a MR spectroscopic image has to be 2-4 cm^3 to acquire proton NMR signals of metabolites with adequate signal-to-noise ratio. Only metabolites with concentrations higher than about 0.1 mM are visible in the MR spectrum. A spin-echo based pulse sequence is often exploited during MRSI in combination with additional pulses to suppress the large proton signal of water and possible fat that would spoil the spectrum[5]. Proton signals of N-acetyl-aspartate (NAA), phosphocreatine and creatine combined (Cr), choline-containing compounds (Cho), and lactate (Lac) can be observed in spatially resolved MR spectra. The metabolite levels in tumor and normal brain tissue are quite different. For example, the concentration of NAA, a neural marker, is low in tumor tissue because of the loss of neuronal cells in tumor tissue.

Cho is elevated in tumor tissue because of the increase in membrane synthesis and degradation. The Cr signal is reduced or even absent in tumor spectra because of the reduced energy metabolism in tumor tissue. Lac is usually observed in tumor spectra because it indirectly marks abnormal glycolysis. Thus the levels of various metabolites, reflected by the signal intensities in the MR spectra, are potential markers for the tumor identity and tumor grade.



However, the observed signal intensities in the MR spectra may not directly reflect the relative concentrations of the metabolites present in the spectroscopic voxels. That is, different combinations of TE and TR yield different spectral profiles of the spectroscopic voxels because the various metabolites have different T_1 and T_2 relaxation times. Only MR spectra of voxels acquired at identical combinations of TE and TR can be compared for tissue characterization. Usually, short echo-time or long echo-time spectra are acquired.

The acquisition of short echo time spectra reduces the signal loss due to T_2 relaxation. This means that spectra with higher signal-to-noise are obtained and additional resonance peaks of metabolites with short T_2 are detected (such as myo-inositol, glutamate, and glutamine). On the contrary, broad background signal appears in the spectra from macromolecules with short T2. Although short echo time spectra provide more biochemical information, they are more difficult to analyze and to interpret due to their contamination with broad peaks of short T_2 macromolecules.

Diagnosing brain tumors

MRI and MRSI will irrevocably improve patient diagnosis and patient treatment if viable tumor tissue can be distinguished in the images from other abnormal non-malignant tissue types within the tumor bed, like necrosis, edema, and possibly radiation necrosis and scar tissue. The delineation of viable tumor will facilitate noninvasive diagnosis, or will at least guide the sampling of biopsies to regions of viable tumor. Moreover, the delineation of viable tumor is very important for studying the effect of chemotherapy or radiation treatment in time as well as for monitoring patients after surgery. Additionally, it helps the radiologist in surgery planning, treatment planning, and in the development and evaluation of new treatment strategies.

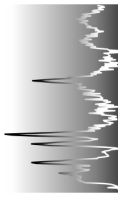
The standard ρ-, T_1-, and T_2-weighted MR images hardly allow proper tissue discrimination within the region affected by the tumor. Contrast agent administration enhances the lesion in the MR image where the blood-brain barrier is defect. This barrier may be variably affected and the extent of tumor tissue may be under- or overestimated in this way. Diffuse regions of abnormal tissue outside the enhanced region may still be seen. Moreover, combined MRI/MRSI studies have revealed that the contrast-enhanced lesion is much smaller than the region of abnormal metabolism[6]. MRSI studies have shown that metabolite levels are highly variable for individual tumors: large standard deviations have been found in metabolite levels of one tumor type and substantial overlap of metabolite levels between different tumor grades[6]. Other studies showed clear differences between necrosis and other abnormal tissue, although these studies are limited by a low spatial resolution[6]. These findings indicate that MRSI and MRI are complementary techniques. Chemometrics should facilitate the fusion of morphologic (MRI) and metabolic (MRSI) data to improve the delineation of brain tumors.

The data analysis methodology must quickly process the large stream of MRI and MRSI data to ensure continuation in the clinical workup of patients. Moreover, it must cope with specific difficulties encountered in the MR data. Due to time constraints, spectroscopic images of low signal-to-noise ratio are acquired. Furthermore, MR data is gathered under in vivo conditions. A patient can move during acquisition and the resulting artifacts may complicate the data analysis. In vivo examinations also hamper any form of "sample preparation" to facilitate the data analysis. Therefore, MR examinations of brain tumor patients are to a large extent at the mercy of proper data processing

A chemometric methodology will only be accepted in the clinical environment if it produces reliable results with minimal user-subjective interaction. Laboratory assistants are usually unfamiliar with sophisticated data analysis techniques. The simplicity and transparency of the data analysis methodology is crucial for its implementation in the hospital.

Processing MR images

Image registration. Image registration, also called image matching, compensates for patient movement between the acquisition of different MR images during the course of a MR examination. The images to be registered may also be from serial MR examinations of a single patient. Serial MR examinations are performed to study for instance tumor growth/shrinkage after radiation treatment. A suitable registration algorithm should then be able to quickly detect the same tissue slice as the one previously measured.



Precautions are taken to restrict the movement of the head during a MR examination of a brain tumor patient. Possible patient movement only results in small shifts between different images of the patient, which can be corrected for by a simple cross-correlation method[7]. However, a disadvantage of cross-correlation methods is their sensitivity to intensity differences in different contrast images. In serial brain tumor MR examinations, feature-based registration methods can help the radiologist in finding the target position of the tissue slice. These methods can perform the alignment of images by means of translation, rotation, and uniform scaling transformations. On beforehand, proper features are sought in the image for matching which are insensitive to changes in tissue and acquisition. These registration points are, for example, edges and corners in normal brain tissue regions. Then a suitable algorithm, like the Procrustes algorithm, can be used to find the least-squares solution by minimizing the distance among all paired points in the two images[8]. The execution time of the algorithm linearly decreases with the number of registration points used. Multiscale approaches using wavelets have been proposed to increase the computation speed of registration algorithms. Image representations on different spatial resolutions allow the sampling of fewer registration points sufficient for adequate registration[9].

Removal of non-brain tissue. Image segmentation of brain tissue is simplified if non-brain material such as bone, skull, face muscles, and air is removed from the images. The removal of these non-brain materials can simply be achieved by drawing a region of interest in the image containing only brain tissue. A simple algorithm for the automatic removal of skull tissue and air is presented in[10]. The method is based on the construction of a radius image of the r-weighted image. The ρ-weighted image is used because it clearly shows the transition between brain tissue and skull. The ρ-weighted image can subsequently be used as a mask image to remove skull tissue and air from

other registered contrast images. Another accepted method for the (semi)automatic removal of the skull is based on a multiresolution algorithm[11, 12].

Image filtering. Image filtering is particularly applied in brain MR imaging of thin tissue slices (thinner than 1 cm). A popular MR image filter is the nonlinear anisotropic diffusion filter[13, 14]. This filter improves the signal-to-noise ratio without blurring of the fine structural details in the image.

Field inhomogeneity correction. Magnetic field inhomogeneities decrease the digitized signal amplitude (pixel intensity) of a voxel in a MR image. An inhomogeneous static magnetic field (B_0) of the scanner broadens the water resonance peak because the resonance frequency is proportional to the strength of the applied magnetic field. This results in a decrease in pixel intensity. Usually, the spatial inhomogeneity of this field is negligibly small.

However, the spatial inhomogeneity of the rf magnetic field (B_1) produced by the coil surrounding the head may be significant. The coil is used to excite the proton spins and to receive the proton signal of a selected tissue slice. Regional field strength differences inside the coil are caused by inhomogeneities in rf coil sensitivity and rf transmission. The non-uniform excitation and reception profiles of the coil yield non-uniform intensity variations across the image. The image non-uniformity depends on the coil design, slice orientation, pulse sequence, and patient. It appears as a low spatial frequency component in the image and is commonly removed by two-dimensional low-pass filtering.

An experimental method to correct for image non-uniformity is to image a uniform sample with the head coil[15]. The correction coefficients for producing uniform intensity images of the sample are stored and are subsequently applied to the brain MR images. An important disadvantage of this method is that the rf penetration through the sample is not the same as through the patient being imaged. A promising method for inhomogeneity correction is based on the acquisition of two identical images under different flip angle[16]. However, this method demands the acquisition of a second image, which takes some valuable acquisition time.

Multivariate image segmentation. A multivariate image consists of a series of MR images of the same tissue slice, e.g., ρ-weighted, T_1-weighted, and T_2-weighted MR images. Segmentation means dividing a multivariate image into regions of different tissue types. Multivariate image segmentation

starts after removal of non-brain tissue and, if necessary, after image registration, field inhomogeneity correction, and image noise removal.

The simplest use of multivariate image segmentation is to combine two registered MR images. For each pixel, the intensity in one image is plotted against the intensity in the other image to define a two-dimensional (2D) variable or feature space. Image segmentation is achieved by defining regions in the 2D feature space corresponding to the different tissue types. In case three registered images are involved, a 3D feature space can be defined in which the different tissue types are located at different positions. This approach becomes, however, more difficult if more than three registered images need to be combined. The lack of visual inspection complicates the isolation of different clusters, corresponding to different tissue types, in the multidimensional feature space. Pattern recognition techniques have been exploited to deal with this problem. Pattern recognition is an important area of chemometrics.

Principal Component Analysis (PCA) is the most popular technique in the field of chemometrics, which has been utilized in the visualization of multivariate image data[17]. PCA transforms a set of registered MR images to a new set of abstract images, called the PCA score images, in such a way that the image contrast information is condensed into the first few PCA score images[18]. By defining a 2D feature space defined by the first two PCA score images, an expert can delineate the different tissue regions in this 2D plot based on 80-90% of the information present in the set of MR images.

However, due to possible field inhomogeneities and the partial volume effect, i.e., the mixing effect of multiple tissue types within one voxel, no distinct groups of pixels appear in the 2D PCA score plot. In fact, no dedicated technique to date is able to reduce the dimension of multivariate MR images in such a way that distinct tissue clusters can be visualized. Therefore, supervised or semi-supervised pattern recognition methods, such as discriminant analysis, neural networks, fuzzy clustering methods, and k-nearest neighbors have been applied to the segmentation of multivariate images. However, these methods may also produce misleading segmentation results because of the subjectivity of the user input. Therefore, PCA is still a valuable tool for visualizing and exploring multivariate images.

Linear discriminant analysis (LDA) is the best-known technique for supervised pattern recognition. LDA utilizes a subset of pixels of known tissue type, a training set, to calculate a classification model. It maximizes the ratio of

the between-class variance and the within-class variance of the tissue classes present in the training set using linear combinations of the original MR images. The calculated discriminant axes (lds), linear combinations of the MR images, discriminate between the tissue classes defined in the training set. The first discriminant axis (ld1) has the largest discriminative power, the next discriminant axes show less discriminative power with increasing ld number. The use of LDA in multivariate image segmentation is illustrated in figure 2.2.

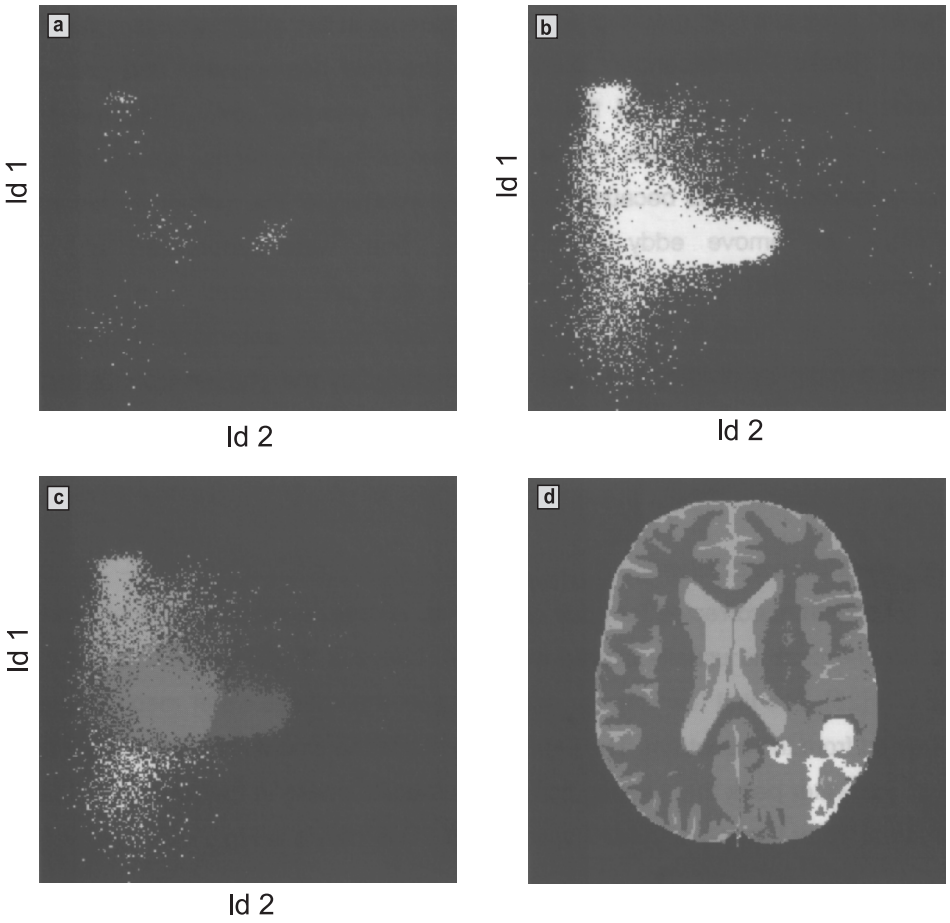


Figure 2.2. LDA applied to the segmentation of a multivariate image. *(a) Discriminant scores of the training pixels for white matter, gray matter, cerebrospinal fluid, and tumor in the ld1-ld2 score plot. (b) Discriminant scores of all pixels in the same ld1-ld2 score plot. (c) Classification of all pixels based on the Euclidean distance: the shortest distance to a class center determines the class membership of the pixel. (d) Classification of all pixels shown in the corresponding image domain. The tissue classes are color-coded: white is tumor, light gray is cerebrospinal fluid, gray is gray matter, and dark gray is white matter.*

As in any supervised or semi-supervised segmentation technique, the training pixels need to be selected by a human interpreter. Different operators will select different training sets (inter-observer variations) at different times (intra-observer variations), resulting in different segmented images or tissue maps. A radiologist has to assess the quality of these calculated tissue maps because there is no definitive method (ground truth) that can verify the segmentation results, which is a fundamental problem in medical image processing. Therefore, segmentation performance is usually determined by the stability of the applied method, i.e., its sensitivity to operator input, and not by the accuracy of tissue volume estimation. Several studies have shown that the segmentation of brain tumors strongly depends on method and training data [19-21]. The major challenge is to develop novel multivariate segmentation techniques that minimize operator input but are robust enough to withstand small variations in input.

Processing MR spectroscopic images

Eddy current correction. Switching magnetic field gradients induces small currents in the magnet system, called eddy currents. These time dependent currents disturb the field homogeneity and cause time dependent frequency shifts of the resonances in the selected voxel. This results in a distortion of the spectrum after Fourier transformation of the time signal (echo). Unlike MRI, MRSI suffers from eddy currents because of the low signal intensity of the metabolite resonances. Therefore, to remove eddy current distortions, both water-suppressed and water-unsuppressed MRSI images of a selected tissue slice are acquired in a single examination, as shown in figure 2.3a.

For each voxel, eddy current correction is accomplished in the time domain by dividing the water-suppressed signal by the phase factor of the water signal for each data point (the influence of metabolite protons in the water-unsuppressed MRS image is negligible)[22]. Figure 2.3b demonstrates a spectrum corrected for eddy currents that is aligned in phase and in spectral position to the water resonance peak within the voxel.

Residual water filtering. The eddy current corrected MR spectrum in figure 2.3b still contains a residual water peak. Filtering of this intense peak is usually done with a Hankel Lanczos Singular Value Decomposition (HLSVD)[23]. The first few singular values account for the intense water resonance and are subtracted from the spectrum. The algorithm is applied in

the time domain to be able to remove the long tails of the water peak overlapping with some metabolite peaks in the spectrum. The result of residual water filtering of the MR spectrum in figure 2.3b is shown in figure 2.3c.

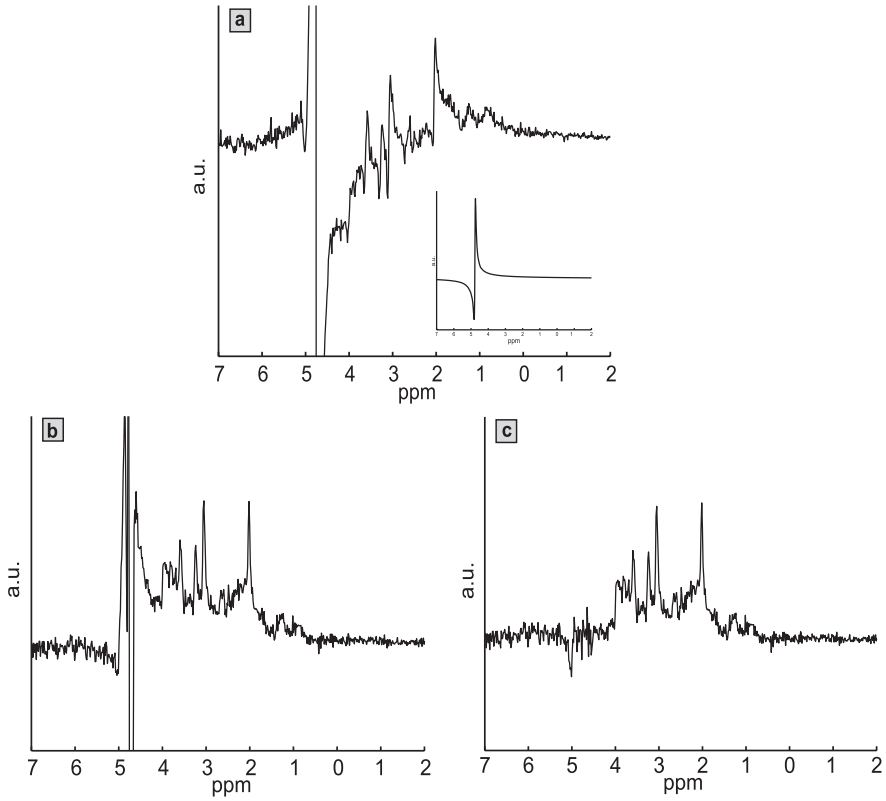
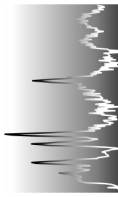


Figure 2.3. *(a) water-suppressed in vivo MR spectrum acquired from a voxel in the human brain. (inset) Water-unsuppressed spectrum from the same voxel. Water suppression is accomplished by using additional rf pulses to saturate the water resonance. (b) After eddy current correction of (a), demonstrating that the metabolite peaks are phased with the phase of the water peak. (c) After filtering the residual water peak at 4.7 ppm of (b).*

Phase and frequency shift correction. Small phase differences between metabolite peaks in the MR spectrum remain after eddy current correction. These remaining phase shifts are frequency-dependent (first-order) phase shifts. Phase correction is performed to obtain complete absorption-mode spectra, which simplify interpretation and analysis of the spectra. MR instruments are standard equipped with interactive phase-correction facilities. However, fast and automatic phasing is desirable for processing the large number of spectra and for obtaining unbiased results.

Frequency shifts, or peak shifts between spectra of different voxels may also distort further analysis of the spectra. Although frequency shifts induced by the spatial inhomogeneity of the rf field are removed after eddy current correction, they may be induced by patient movement between the acquisition of water-unsuppressed and water-suppressed MRSI data. Patient movement changes the coil loading and as a result changes (deteriorates) the spatial homogeneity of the rf field inside the coil. A novel method has recently been developed for removing quickly phase shifts and frequency shifts across a large series of single resonance peaks[24]. The method performs well on noisy and distorted MR spectra acquired under in vivo conditions.



Pattern recognition. After preprocessing, MR spectra can be classified according to their tissue identity. It is impossible to categorize the processed MR spectra with respect to tissue identity by visual inspection of the data, which makes pattern recognition methods indispensable. These methods cluster the spectra on their mutual characteristics.

Selecting spectral features that provide the best discrimination are important considerations when designing a pattern recognition model. For the pattern recognition of brain MR spectra, the estimated levels of metabolites are used rather than the complete spectrum. The relative levels of NAA, Cho, and Cr are derived from long echo time spectra by curve fitting routines. These estimated levels considerably reduce the dimension of the input data of the pattern recognizer. Moreover, spectral noise is omitted in the pattern recognition.

Metabolite levels from short echo time spectra are much harder to estimate. Short echo time spectra are contaminated with broad peaks of non-relevant macromolecules which distort the classification performance. The fitting of short echo time spectra is usually done in the time domain. An important advantage of time domain fitting is that complicated model functions can be used to improve the fitting of short echo time spectra. For example, the useful Voigt function can only be approximated in the frequency domain or must be calculated by numerical integration[25]. Nonlinear iterative methods, such as the Levenberg-Marquardt algorithm[26], are preferably utilized in spectral fitting. These methods allow any mathematical form of the model function, like the Voigt function.

Moreover, prior knowledge of, for example spectral position and relative amplitudes can be incorporated in the model function and improve the quantification of resonance peaks[27]. Recently, novel quantitation methods have been developed which operate on complete series of MR spectra instead of on each separately[24,28-30]. These have been proposed for the fast and accurate quantitation of resonances in series of relatively simple in vivo MR spectra.

Figure 2.4a demonstrates the pattern recognition of a spectroscopic image of a patient with a brain tumor. Two of the MR spectra acquired from the patients brain are plotted in figure 2.4b. The six resonance peaks labeled in the spectra were quantified and subsequently normalized to the unsuppressed water signal within the voxel. These normalized quantified peaks define the input vector for the pattern recognizer.

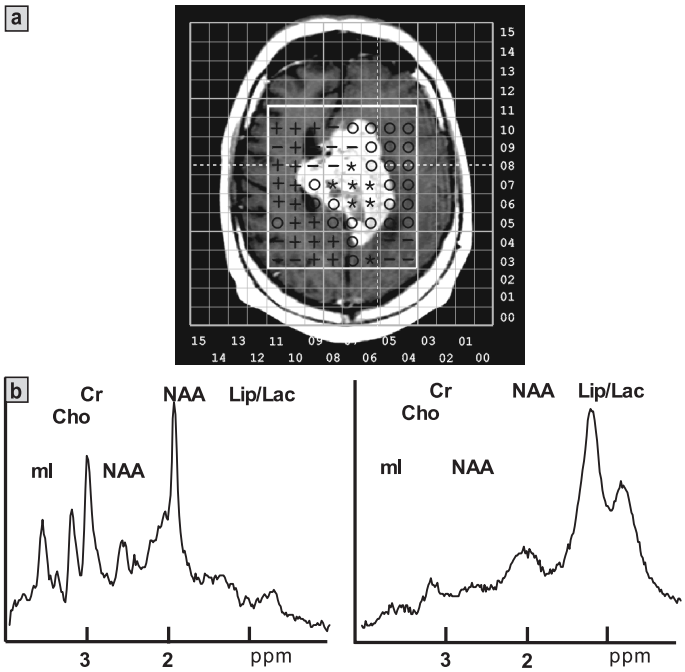


Figure 2.4. *(a) Background: Gd-enhanced T_1-weighted MR image of a brain tumor. Foreground: Grid of spectroscopic voxels. The symbols indicate the group membership of each spectrum, as determined with hierarchical cluster analysis. (b) Two of the MR spectra of the spectroscopic image, belonging to two different groups. The left one originates from a healthy region, the right one from a malignant region.*

The MR spectra, as represented by the aforementioned input vectors, could roughly be divided into four different groups, which was performed by means of hierarchical clustering. The result of the hierarchical clustering analysis is plotted as a dendrogram in figure 2.5. Hierarchical clustering merges the individual samples into groups, which are subsequently merged until only one group remains. The merging of the samples is based on the distances between the input vectors, which can be the Euclidean distance. Selection of the number of groups is performed by setting a threshold which cuts the dendrogram.

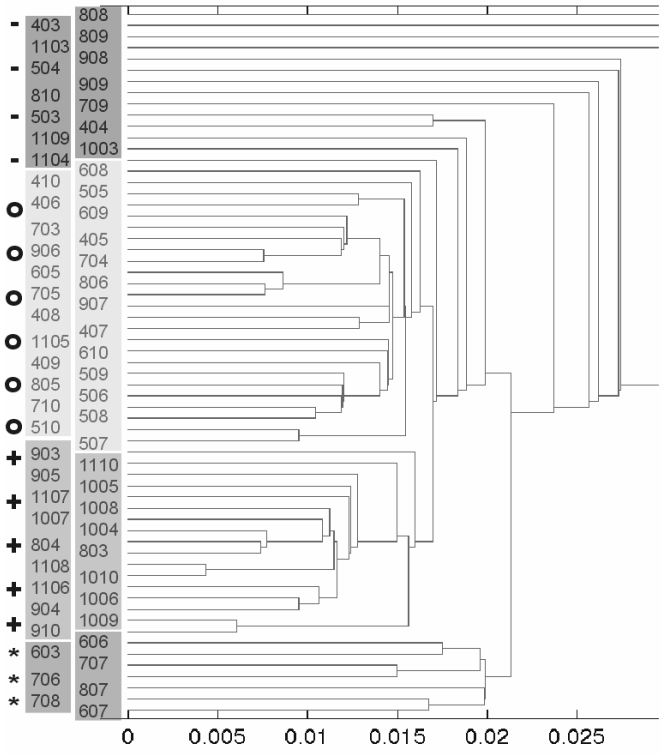


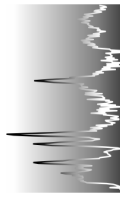
Figure 2.5. Hierarchical clustering of MR data. *The data of the patient depicted in figure 2.4a. is used. The symbols on the left of the figure correspond with the symbols used in figure 2.4a. The horizontal axis represents the Euclidean distance.*

PCA and LDA can also be used to cluster the MR spectra of different patients with brain tumors. These methods also use the six quantified resonances as input. With PCA, a set of linear combinations of the quantified peaks, called the principal components, is constructed in such a way that as much as possible variation is squeezed in as few as possible new variables (PCs). The PCA score plot defined by the first two PCs yields a visualization of the spectra, which is based on a large part of the total information. With LDA a set of linear combinations of the quantified peaks, called the discriminant functions, is constructed which have the highest possible discrimination power among the classes of spectra defined on beforehand. The LDA model is biased by the labeling of the spectra by the investigator. The application of LDA to MRSI data is problematic because accurate histological characterization of all spectroscopic voxels is often impossible. Thus, overall, a reliable discrimination among brain tumors based on MRSI has yet to be accomplished, although progress has been made[31]. Some clinical trial studies involving multiple sites have attempted to combine MRSI and MRI data to improve tumor discrimination.

Prospects of combining MR images

On one hand, MR spectroscopic imaging provides metabolic information on a low spatial resolution; on the other hand, MR imaging provides morphological information on a high spatial resolution. The combination of MR imaging and MR spectroscopic imaging improves the MR characterization of brain tumors. The levels of metabolites differ in different parts of the brain. For example, the MR spectra of normal (left) and malignant tumor (right) in figure 2.4b clearly show differences. In addition, several studies have demonstrated that levels of NAA, Cr, and Cho differ between white matter and gray matter and may be affected by the contribution of cerebrospinal fluid[32-34]. Therefore, to classify the spectra accurately, the MR spectra must be normalized to the tissue composition within the spectroscopic voxel.

Classifying the MR spectra on the basis of the estimated metabolite levels and the fraction of white matter, gray matter, cerebrospinal fluid, viable tumor, and necrosis within the spectroscopic voxel may be the best approach. In this way image and spectroscopic data are merged into one classifier or pattern recognizer. Several groups are developing an overall classifier of MR image and spectroscopic data[35]. Such a classifier is also being developed at the University Medical Center Nijmegen, the Netherlands.



The measurement protocol involves the acquisition of the standard ρ-weighted, T_1-weighted, and T_2-weighted spin echo images, a T_1-weighted spin echo image after contrast agent administration, a so-called relative regional cerebral blood volume (rCBV) image, and a short echo time spectroscopic image. The five MR images define a multispectral or multivariate image in which each pixel (image element) is characterized by its pixel intensities in the five MR images.

The multivariate image will be segmented into regions of white matter, gray matter, cerebrospinal fluid, viable tumor, and necrosis. Hopefully, patients will benefit from advanced evaluation and treatment by combining the segmented image with the spectroscopic image to improve the demarcation of the brain tumors from healthy tissue and the delineation of tissue heterogeneity within the tumor bed.

References

1 DL Massart, BGM Vandeginste, LMC Buydens, S de Jong, PJ Lewi, J Smeyers-Verbeke. *Handbook of Chemometrics and Qualimetrics: Part A*; Elsevier Science Publishers: Amsterdam. Data Handling in Science and Technology, vol 20A, 1998.

2 BGM Vandeginste, DL Massart, LMC Buydens, S de Jong, PJ Lewi, J Smeyers-Verbeke. *Handbook of Chemometrics and Qualimetrics: Part B*; Elsevier Science Publishers: Amsterdam. Data Handling in Science and Technology, vol 20B, 1998.

3 M Otto. *Chemometrics, Statistics and Computer applications in Analytical Chemistry*; Wiley-VCH: Weinheim, 1999.

4 KR Beebe, MB Seasholtz. *Chemometrics A Practical Guide*; John Wiley and Sons: New York, 1998.

[5] DG Gadian. *NMR and its applications to living systems*; Oxford University Press: Oxford, 1995.

[6] SJ Nelson, DB Vigneron, WP Dillon. Serial evaluation of patients with brain tumors using volume MRI and 3D ^{1}H MRSI, NMR in Biomedicine 1999; **12**: 123-138.

[7] WK Pratt. *Digital image processing*; John Wiley & Sons: New York, 526-566, 1978.

[8] EC Evans, DL Collins, P Neelin, D MacDonald, M Kamber, TS Marret. *Three-dimensional correlative imaging: Applications in human brain imaging*; Orlando, FL: Academic Press, Functional Neuroimaging: Technical Foundations (Eds RW Thatcher, M Hallett, T Zeffiro, ER John, M Huerta), 1994, 145-161.

[9] ME Alexander, RL Somorjai. The registration of MR images using multiscale robust methods, Magn Reson Imaging 1996; **14**: 453-468.

[10] Z Liang, D Wang, J Ye, D Harrington. Development of automatic techniques for segmentation of brain tissues from multispectral images, IEEE Nuclear Science Symposium 1995; **3**: 1453-1456.

[11] H Soltanian-Zadeh, JP Windham, F Chen. Automated contour extraction for medical image registration and segmentation, Proc SPIE Medical Imaging 1995: Image Processing Conference, San Diego, 1995

[12] H Soltanian-Zadeh, JP Windham, DJ Peck, T Mikkelsen. Feature space analysis of MRI, Magn Reson Medicine 1998; **40**: 443-453.

[13] P Perona, J Malik. Scale space and edge detection using anisotropic diffusion, IEEE Trans Patt Anal Machine Intell 1990; **12**: 629-639.

[14] G Gerig, O Kubler, R Kikinis, FA Jolesz. Nonlinear anisotropic filtering of MRI data, IEEE Trans Med Imaging 1992; **11**: 221-232.

[15] FB Mohamed, S Vinitsky, C Gonzalez, S Faro, C Burnett, HV Ortega, T Iwanaga. Image nonuniformity correction in high field (15T) MRI, Proc IEEE Eng Med Bio 1995; **17**: 36-37.

[16] H Mihara, N Iriguchi, S Ueno. A method of RF inhomogeneity correction in MR imaging, Magnetic Resonance Materials in Physics, Biology and Medicine 1998; **7**: 115-120.

[17] P Geladi, H Grahn. *Multivariate image analysis*; John Wiley & Sons: Chichester, 1996

[18] P Geladi, H Isaksson, L Lindqvist, S Wold, K Esbensen. Principal component analysis of multivariate images, Chemom Intell Lab Syst 1989; **5**: 209-220.

[19] LP Clarke, RP Velthuizen, S Phuphanich, et al. MRI: stability of three supervised segmentation techniques Magn Reson Imaging 1993; **11**: 95-106.

[20] M Vaidyanathan, LP Clarke, RP Velthuizen, et al. Comparison of supervised MRI segmentation methods for tumor volume determination during therapy, Magn Reson Imaging 1995; **13**: 719-728.

[21] M Vaidyanathan, LP Clarke, LO Hall, et al. Monitoring brain tumor response to therapy using MRI segmentation, Magn Reson Imaging 1997; **15**: 323-334.

[22] U Klose. In vivo proton spectroscopy in presence of eddy currents, Magn Reson Medicine 1990; **14**: 26-30.

[23] R de Beer, D van Ormondt. Analysis of NMR data using time domain fitting procedures; Springer, Berlin, NMR Basic Principles and Progress 1992; **26**: 201-248.

[24] H Witjes, WJ Melssen, HJA in 't Zandt, M van der Graaf, A Heerschap, LMC Buydens. Automatic correction for phase shifts, frequency shifts, and lineshape distortions across a series of single resonance lines in large spectral data sets, J Magn Reson 2000; **144**: 35-44.

[25] I Marshall, J Higinbothan, S Bruce, A Freise. Use of Voigt lineshape for quantification of in-vivo ^{1}H spectra, Magn Reson Medicine 1997; **37**: 651-657.

[26] DW Marquardt. An algorithm for least-squares estimations of non-linear parameters, J Soc Ind Appl Math 1963; **11**: 431-441.

[27] R Bartha, DJ Drost, PC Williamson. Factors affecting the quantification of short echo in-vivo ^{1}H MR spectra: prior knowledge, peak elimination, and filtering, NMR in Biomedicine 1999; **12**: 205-216.

[28] R Stoyanova, AC Kuesel, TR Brown. Application of principal-component analysis for NMR spectral quantitation, J Magn Reson A 1995; **115**: 265-269.

[29] MA Elliot, GA Walter, A Swift, K Vandenborne, JC Schotland, JS Leigh. Spectral quantitation by principal component analysis using complex singular value decomposition, Magn Reson Medicine 1999; **41**: 450-455.

[30] Y Wang, S van Huffel, L Vanhamme, P van Hecke. Magnetic resonance spectroscopic quantitation via complex principal component analysis and Hankel total least squares based methods, SISTA report 2000-18.

[31] MC Preul, Z Caramanos, R Leblanc, JG Villemure, DL Arnold. Using pattern analysis of in vivo proton MRSI data to improve the diagnosis and surgical management of patients with brain tumors, NMR in Biomedicine 1998; **11**: 192-200.

[32] Y Wang, S-J Li. Differentiation of metabolic concentrations between gray matter and white matter of human brain by in vivo ^{1}H Magnetic Resonance Spectroscopy, Magn Reson Medicine 1998; **39**: 28-33.

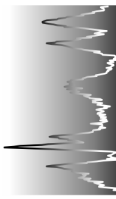
[33] P Pouwels, J Frahm. Regional metabolite concentrations in human brain as determined by quantitative localized proton MRS, Magn Reson Medicine 1998; **39**: 53-60.

[34] S M Noworolski, SJ Nelson, RG Henry, et al. High spatial resolution ^{1}H-MRSI and segmented MRI of cortical gray matter and subcortical white matter in three regions of the human brain, Magn Reson Medicine 1999; **41**: 21-29.

[35] HP Hetherington, JW Pan, GF Mason, et al. Quantitative ^{1}H spectroscopic imaging of human brain at 41 T using image segmentation, Magn Reson Medicine 1996; **36**: 21-29.

Chapter 3

Automated correction of unwanted phase jumps in reference signals which corrupt MRSI spectra after Eddy current correction‡



A.W. Simonetti[♣]*, W.J. Melssen*[♣]*, M. van der Graaf*[♦]*, A. Heerschap*[♦] *and L.M.C. Buydens*[♣]

Laboratory for Analytical Chemistry, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

Department of Radiology, University Medical Center Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

‡ Published in Journal of Magnetic Resonance 2002; **159**, p 151-157

Abstract

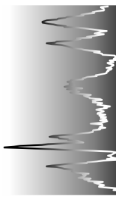
A commonly applied step in the postprocessing of gradient localized proton MR Spectroscopy, is correction for eddy current effects using the water signal as a reference. However, this method can degrade some of the metabolite signals, in particular if applied on proton MR Spectroscopic Imaging data. This artifact arises from the water reference signal in the presence of a second signal which resonates close to the main water resonance. The interference of both resonances will introduce jumps in the phase of the reference time domain signal. Using this phase for eddy current correction will result in a ringing artifact in the frequency domain of the metabolite signal over the whole frequency range.

We propose a moving window correction algorithm, which screens the phase of reference signals and removes phase jumps in time domain caused by interference of signals from multiple spin systems. The phase jumps may be abrupt or gradually distributed over several time data points. Because the correction algorithm only corrects time data points which contain phase jumps, the phase is minimally disrupted. Furthermore, the algorithm is automated for large datasets, correcting only those water reference signals which are corrupted. After correction of the corrupted reference signals, normal eddy current correction may be performed. The algorithm is compared with a method which uses a low-pass filter and tested on simulated data as well as on in vivo proton spectroscopic imaging data from a healthy volunteer and from patients with a brain tumor.

Keywords: Magnetic Resonance Spectroscopic Imaging, eddy current correction, data processing, reference deconvolution, artifact.

Introduction

In proton magnetic resonance spectroscopy, eddy current correction (ECC)[1-5] is a standard correction method[6] which requires a metabolite signal as well as a reference signal. In in vivo MR studies, the reference signal normally consists of the unsuppressed water signal. Theoretically, ECC corrects for time dependent eddy current phase effects in the free induction decay (FID). Besides, it resets the zero-order phase and resets the water frequency to 0 Hz. However, ECC may lead to a severe ringing artifact in the metabolite spectrum[7-9] if the envelope of the water reference time domain signal crosses zero, causing the phase of the water reference signal to flip 180 degrees at a specific time point. The zero crossing of the envelope may occur because the single water reference peak has acquired some doublet character because of poor shimming or due to susceptibility effects[7,9]. Generally, these artifacts appear more frequently in magnetic resonance spectroscopic imaging (MRSI) data than in single voxel data because of the much larger volume of interest. In addition, signals with slightly different water frequencies originating from different voxels in the MRSI dataset may be mixed as a result of the point spread function[10]. The eddy current division will introduce the 180 degree phase jumps into the metabolite time domain signal, causing severe ringing in the frequency domain after Fourier transformation.



A solution to this problem has been presented by Wild[7], who used low-pass filtering of the total phase of the water reference time domain signal to remove unexpected phase jumps. They tested the efficiency of two low-pass filters, an infinite impulse response Butterworth filter and a finite impulse response filter. The Butterworth filter was found to be more effective.

This paper also deals with the implementation of a correction algorithm which detects phase jumps in the phase of reference time domain signals, but instead of filtering the whole phase our algorithm has a local nature. It searches for specific features in the magnitude and the phase of the reference time domain signal, and only corrects time points when specific conditions are met. The advantages of our method are; (a) only the time points at a phase jump are altered; (b) reference signals which do not contain phase jumps remain unchanged and therefore the correction algorithm can be applied on large datasets like MRSI data, and (c) the correction algorithm identifies abrupt changes as well as more gradual distributed transitions, which also induce ringing in the frequency domain of a metabolite signal.

In order to validate the correction algorithm, we tested it on simulated data, compared it with the filter described by Wild[7] and used it to screen brain MRSI datasets of a volunteer and several patients with a brain tumor. In 93% of the cases the unwanted phase jumps in the reference time domain signal were identified by our method, resulting in substantial improvement of the metabolite spectrum after ECC.

Theory

Eddy current correction. The theory of ECC has been frequently described in literature[1-5, 8] and will therefore be discussed concisely. To perform ECC, the metabolite time domain signal is point-wise divided by the phase factor of the reference signal in the time domain. Practically, the phase of the reference signal is calculated by taking the arctangent of the quotient of the imaginary and real part of each time point.

The correction only holds when the reference signal contains a single dominant component. If this is not the case, one should use other techniques like described by[9]. However, in the correction of large sets of data, as in MRSI, which only contain few reference signals with some doublet character, ECC is still very effective. Correction of the largest artifact introduced in the metabolite spectrum, i.e. the ringing due to phase jumps in the phase of the reference time domain signal, ensures that the metabolite spectrum can still be of use.

The ringing artifact. The origin of the ringing artifact is explained in figure 3.1. As an example we have simulated a metabolite and two reference signals. The metabolite signal contains two resonances. The first reference signal contains one resonance and the second two. These two resonances are close together, in order to simulate a water reference signal with some doublet character. Information about the signals can be found in table 3.1.

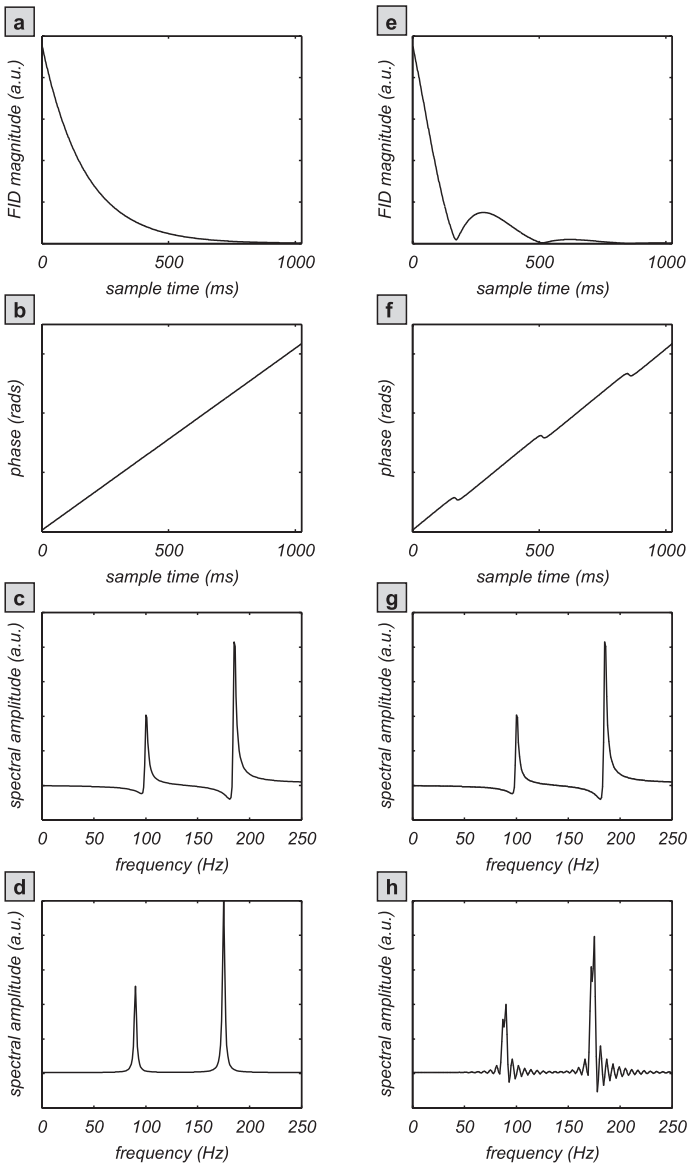


Figure 3.1. Example of ECC and the ringing artifact. *Parameters for the simulated data can be found in table 3.1. (a) Magnitude of the simulated reference time domain signal containing 1 resonance. (b) Phase of this reference signal. (c) Frequency domain of the simulated metabolite spectrum. (d) ECC-corrected metabolite spectrum. The zero-order phase is corrected and the two resonances are shifted 10 Hz to the left. (e) Magnitude of the simulated reference signal containing 2 resonances. (f) Phase of this reference signal. The phase jumps are gradually distributed but will nevertheless introduce ringing in the metabolite spectrum (g) when ECC is performed (h).*

The left column in figure 3.1 gives an example if the reference signal only contains 1 resonance. Figure 3.1a and b show the magnitude and phase of the reference time domain signal. Both magnitude and phase do not show any irregularities. When the phase is used to perform ECC on the metabolite signal (figure 3.1c), this results in a phase corrected spectrum (figure 3.1d) which has shifted 10 Hz to the left due to the offset frequency of the reference to the detector.

In the right column of figure 3.1 the same metabolite spectrum is used (figure 3.1g), but in this situation the reference signal contains two resonances which are only 3 Hz apart. Because of interference, the magnitude (figure 3.1e) of the reference signal shows beats at specific points in time domain. The phase of the reference signal contains phase jumps which are gradually distributed and not exactly 180 degrees, because the amplitudes A1 and A2 are not the same (figure 3.1f). ECC of the metabolite FID with the phase of this reference signal will introduce severe ringing in the spectrum (figure 3.1h). This ringing is asymmetrical around the tops of the resonances, because the phase jumps are distributed over a certain number of points. If the phase jump occurs over one time point the ringing is symmetrical, but in practice this hardly ever occurs.

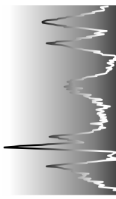
Table 3.1. Parameters for the simulated data presented in figure 3.1[a]

	metabolite signal	**reference signal 1**	**reference signal 2**
ω_1	100	10	10
ω_2	185		13
$d_{1,2}$ **(Hz)**	8	6	6
$\Phi_{1,2}$ **(°)**	35	35	35
A_1 **(a.u.)**	1	95	50
A_2 **(a.u.)**	2		45

[a] *Each signal is constructed using the following equation:*
$FID = \sum_k A_k * \exp(i2\pi / 360\Phi_k) * \exp(-\pi d_k + i2\pi\omega_k)t_n$, *with* A_k *the amplitude,* Φ_k *the phase,* d_k *the damping and* ω_k *the frequency of resonance k.* t_n *is the time instance.*

Outline of the correction algorithm. To remove the jumps in the phase of the reference time domain signal, we have developed a correction algorithm which moves along it, correcting it if certain conditions are met. If one of the conditions fails, the reference signal is not changed. A flow-chart of the steps in the correction algorithm is presented in figure 3.2.

Step a: if the reference signal contains two interfering resonances, its magnitude will contain one or more minima. Normally only one minimum occurs in the signal-containing part of the magnitude of the FID. However, if necessary it is possible to search in a successive way for more minima and investigate if the phase at that point also needs a correction. Each minimum can be found by searching for a change in the sign of the first derivative of the magnitude signal from negative to positive. Reference signals from in vivo experiments normally have a large signal to noise ratio, therefore finding a zero crossing in the first derivative is straightforward. If the signal is noisy, smoothing of the magnitude can be applied.



Step b: a minimum in the magnitude must coincide with a sigmoidal shape in the phase of the reference signal. This shape can be found in the first derivative of the phase, by searching for a successive number of data points which are above or below the baseline.

Step c: the location of the minimum in the magnitude is compared to the location of the sigmoidal shape. They have to be at the same time point. If there is no minimum, no sigmoidal shape or the two do not coincide, then there is no phase jump due to interference and the correction algorithm is aborted.

Step d: to find the beginning and the end of the phase jump, the first derivative of the phase is fitted to a model function (e.g. Gaussian). This can be done in a robust way, because the position and maximum are known and the width is rather small for phase shifts which corrupt the metabolite spectrum after ECC. There is no theoretical basis for fitting to a Gausssian, but in practice this works fine.

Step e: from the fitted first derivative of the phase, the beginning and the end of the part of the phase which needs to be corrected can be found. The first point of this region can be found by subtracting twice the FWHM from the x-value belonging to the maximum of the first derivative of the phase. The last point is found by adding twice the FWHM to the x-value belonging to the maximum. The factor of two ensures that all corrupted points are corrected.

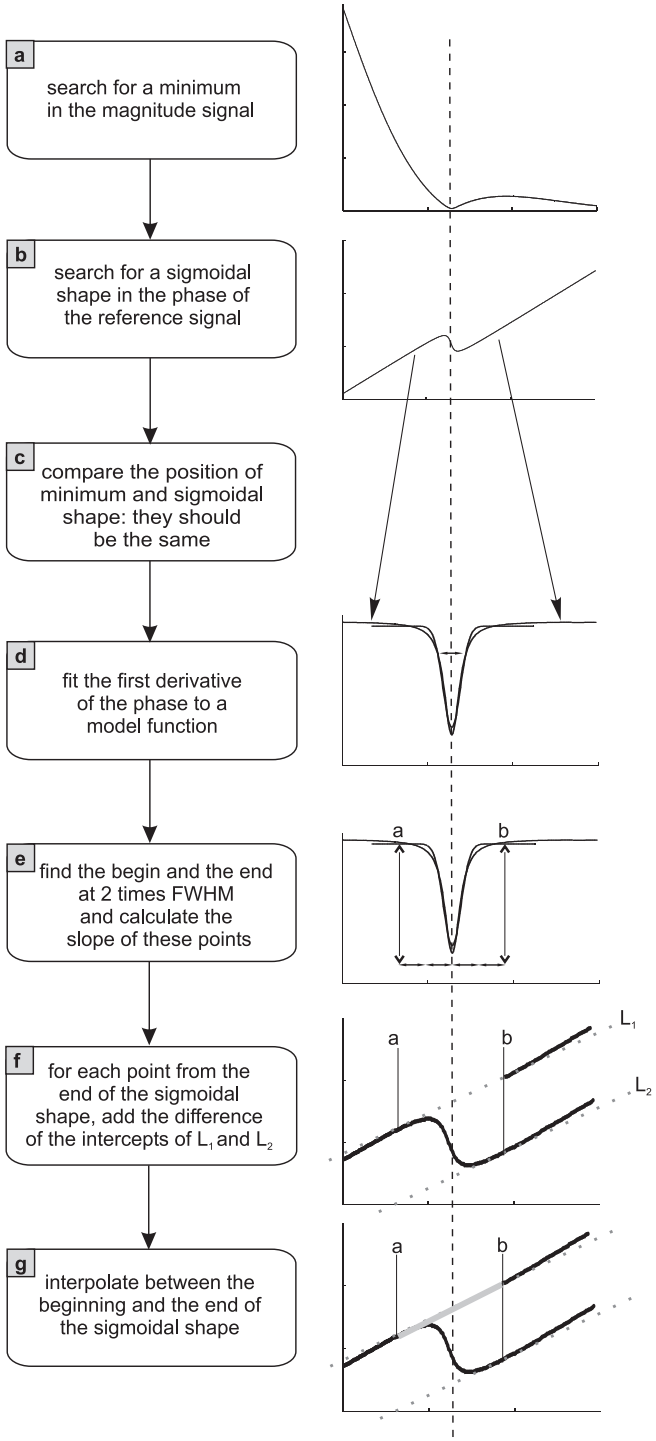


Figure 3.2. Flow-chart of the correction algorithm

Step f: the slope of the phase at the start and end point is set equal and is approximated by taking the offset of the fitted first derivative. This slope is used to plot two lines (L1 and L2) which are parallel and go through the begin and the end point. The part of the phase after the end point is corrected by adding the difference of the intercepts of L1 and L2 for each point.

step g: finally, the part between the begin and the end point is interpolated, resulting in a straight line with the slope of L1 between these points. This part of the phase is approximated and replaces the incorrect sigmoidal shape.

If one of the steps a to d fails, the original phase of the reference signal will be used. This ensures that only appropriate corrections are performed. Although the correction algorithm is assumed to be robust for data acquired using several acquisition protocols (e.g. PRESS, STEAM), some parameters have to be checked and adjusted. The phase of the reference signal may become noisy when the magnitude approaches zero. Correcting it in this part of the signal will become difficult, but is also not necessary because it is more probable that a phase jump in that area is introduced due to the noise in the signal than due to interference of resonances. Therefore we only screen the phase of the reference signal until its magnitude has decreased to 1 percent of its maximum. A second parameter which can be adjusted concerns the maximum width of the fitted model function in step d of the correction algorithm's design. When the width increases, the total number of time data points which are interpolated will also increase. This may prevent a good correction. Furthermore, a large width of the fitted model function indicates a slowly changing phase, which will not introduce ringing. Therefore, a maximum should be set for this parameter. If the fitted model function exceeds this maximum, the correction algorithm is not triggered. For our in vivo data we found a width of 40 ms to be convenient.

Experimental

In order to compare our algorithm with the filter implemented by Wild, we copied his simulated data, which consisted of the metabolite and reference signals listed in table 3.2.

In addition, water suppressed and unsuppressed 2D ^{1}H MRSI datasets of the human brain were acquired from one healthy volunteer as well as from patients with a brain tumor. The measurements were performed on a 1.5 T Siemens Vision whole body MR system, with the following parameters: 16x16 matrix size, STEAM volume pre-selection and outer volume suppression, a repetition time of 2500 ms, an echo time of 20 ms, a slice thickness of 15 mm and a field of view of 200 mm. The number of time points sampled was 1024 with a dwell time of 1 msec.

Table 3.2. Parameters for the simulated data presented in figure 3.3[a]

	metabolite signal	**reference signal**
ω_1 (Hz)	8.1	8.1
ω_2 (Hz)	97.3	16.2
ω_3 (Hz)	110.6	
ω_4 (Hz)	176.0	
d_1 (Hz)	22.2	2.2
$d_{2,3,4}$ (Hz)	3.2	2.2
$\Phi_{1,2,3,4}$ (^{o})	0	0
A_1 (a.u.)	20	1000
$A_{2,3,4}$ (a.u.)	2	1000

[a] *For explanation of the parameters we refer to table 3.1.*

The acquired metabolite- and reference- data were first Fourier transformed in the direction of the spatial axes only. Mild Hamming smoothing was applied prior to Fourier transform to lower the ripple effect[10] introduced by the point spread function.

Only mild smoothing was used to prevent significant broadening of spatial structures. Before ECC, all reference signals within the STEAM box of each MRSI dataset (~80 time domain signals) were automatically screened and corrected for phase jumps -if present- as described. After ECC of all metabolite signals, the residual water in the metabolite signal was removed. This was performed by HLSVD filtering[11] of a region between 4.3 and 5.1 ppm in the spectrum, with 10 singular values. The HLSVD filtering enables a better comparison of the relevant part of the spectrum. Finally, each metabolite time domain signal was Fourier transformed to get 1H MR spectra.

Results and Discussion

The use of our correction algorithm on the simulated data (table 3.2) and the comparison with the one Wild presented is visualized in figure 3.3. The phase of the reference signal, which is plotted in figure 3.3a, shows 8 phase jumps. When this phase is used to perform ECC on the metabolite signal, the result is a heavily corrupted spectrum (figure 3.3b). In figure 3.3c, the corrected phase using our correction algorithm is shown (dotted line). The jumps in the corrected phase are totally removed, except for the two last ones. At that location the amplitude of the FID magnitude is below 1% of its maximum, causing the moving window algorithm to stop. After ECC of the metabolite signal, this results in a virtually uncorrupted spectrum (figure 3.3d). The effect of low-pass filtering the phase of the reference signal using Wild's method is presented in figure 3.3e. The phase steps have been removed effectively (dotted line), although a delay has been introduced and the phase is curved at the first 200 points. The resulting spectrum after ECC (figure 3.3f) is much better than the original, but still some artifacts are visible at the base of the peaks. Wild has shown that mild smoothing of the frequency domain will effectively remove the artifacts at the base.

In *in vivo* MRSI data, the moving window correction algorithm was found to be very effective and robust in correcting data in batch mode. In one volunteer and several patients, the screening resulted in correction of up to 10% of the reference signals. Of this 10%, some of the reference signals are erroneously corrected due to supposed phase jumps in the last (noisy) part of the phase. However, these erroneous corrections are not harmful, because they occur in the noise.

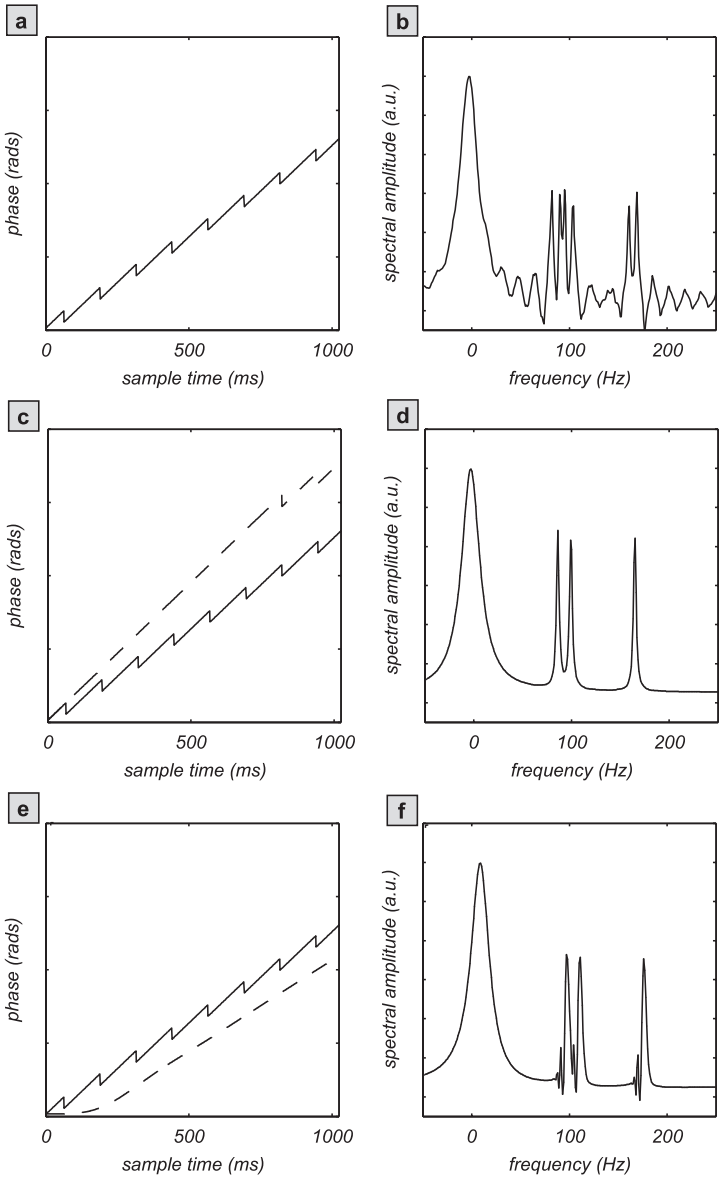


Figure 3.3. Comparison of ECC corrected simulated spectra. *When the non-corrected phase in (a) is used, the metabolite spectrum is heavily corrupted (b). When the phase is corrected with our algorithm (dotted line in c), the metabolite spectrum (d) is virtually uncorrupted. In (f) the metabolite spectrum after correction with the phase filtered by Wild's method (dotted line in e) is shown.*

In figure 3.4 an example is given of one corrected time domain signal from the volunteer. Figure 3.4a and b show the magnitude and phase of the reference signal. Figure 3.4c shows the ECC-corrected metabolite spectrum if the phase shown in figure 3.4b is used. Ringing is especially visible in the residual water signal, but is present over the whole frequency range. In figure 3.4d, the first derivative of the phase is plotted together with its fitted function (dotted). Information from the fit is used to correct the phase of the reference signal as described (figure 3.4e), and this phase is used for ECC of the metabolite signal. The resulting ECC-corrected metabolite spectrum is plotted in figure 3.4f and is virtually free of ringing.

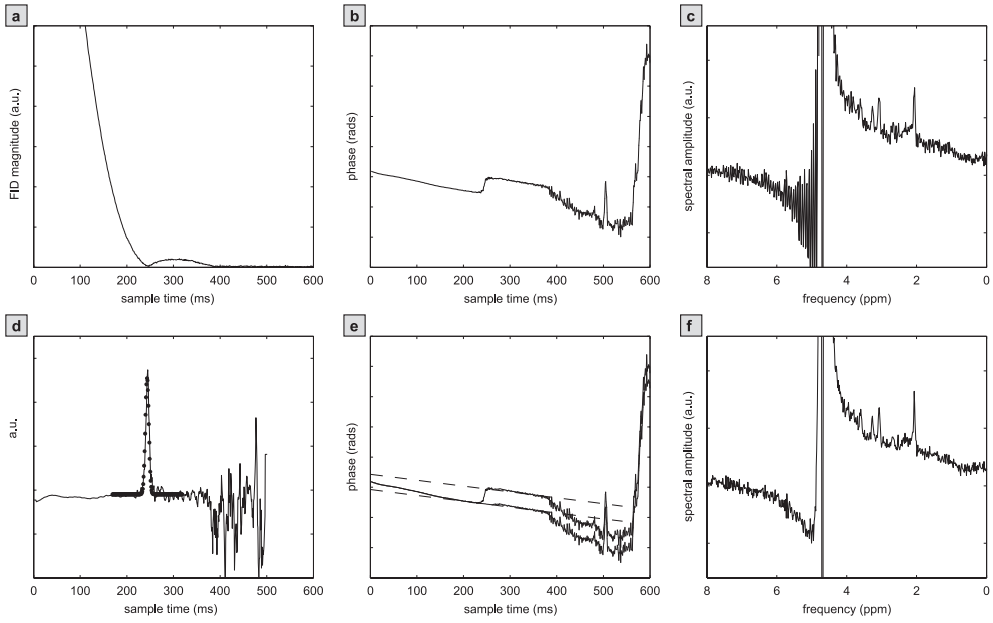


Figure 3.4. Example of the application of the correction algorithm on MRSI data from a volunteer. *(a) Magnitude and (b) phase of the reference time domain signal. (c) The metabolite spectrum after ECC correction with the phase shown in (b). (d) The first derivative of the phase locally fitted to a model function. (e) Correction of the phase as explained in the theory section. f) The metabolite spectrum after ECC correction with the corrected phase.*

Figure 3.5 gives examples of corrected spectra from which the water residual has been removed by HLSVD. The upper spectrum of each subplot has been ECC-corrected with the uncorrected phase of the water reference signal. In the middle of a subplot, spectra are depicted which have been corrected by our method. It is clearly visible that; a) in the uncorrected cases a lot of residual water remains after removal of the water resonance by HLSVD; b) The metabolic resonances in the corrected spectra have narrower lineshapes and have higher amplitudes, since line broadening by ringing is removed, and c) in cases of severe ringing or low signal to noise ratio some resonances are completely flooded in the uncorrected case, thus making it impossible to quantitate these resonances in the frequency domain[12]. Also quantitation in the time domain[13] will be difficult, because the metabolite FID contains an unexpected phase jump. The lower spectrum of each subplot shows the corrected spectrum using Wild's filter. After manually setting the filter parameters, it performs under visual inspection equally well as ours.

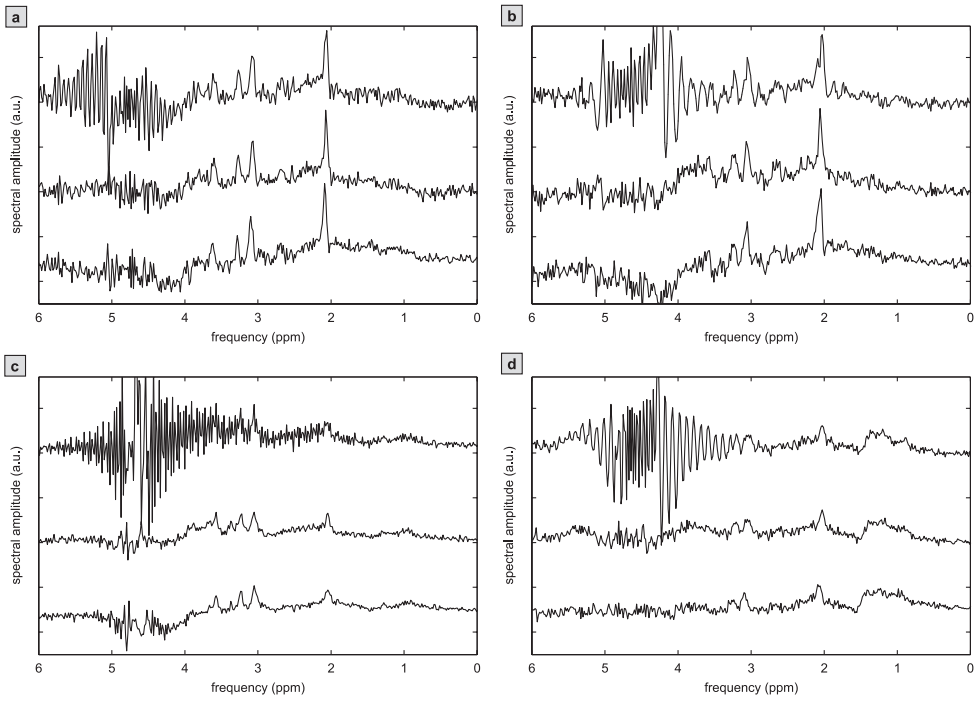
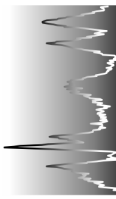


Figure 3.5. Examples of spectra. *(a) from a volunteer and (b, c and d) several patients. The upper spectrum is ECC-corrected without correcting the phase of the reference signal, the middle spectrum is ECC-corrected using the presented correction algorithm, the lower with Wild's filter.*

Robustness of the correction algorithm is based on the fact that reference signals are only corrected when they satisfy three criteria; a) a local minimum in the magnitude reference signal; b) a sigmoidal shape in the phase of the reference signal, and c) the possibility to fit a model function through the first derivative of the phase with a low error and a relatively small width. When one of these criteria is not met, it is assumed that correction is not needed. Next to robustness, these criteria also ensure that only phase jumps caused by two interfering resonances are corrected, therefore the algorithm can be applied on large datasets.

The algorithm also takes the steepness of the phase jump in the reference signal into account; when the discontinuities in the magnitude and phase are large (and ringing is severe), they are easily detected. If the discontinuities are more gradual, they will be more difficult to correct, but this is not a problem because the ringing will be less.



Conclusion

We have presented a method to locally remove jumps in the phase of water reference time domain signals. The removal of these jumps prevents the metabolite signal to become corrupted after ECC by severe ringing, which may completely obscure all resonances. The correction algorithm only corrects specific regions of the phase and is able to do so for sharp as well as gradually distributed transitions, which both can occur if the water reference obtained some doublet character.

In contrast to low-pass filters, our method uses features within the reference time domain signal which are related to the artifact. Using these features ensures a robust correction and enables the application of the algorithm in an automated manner on large datasets like MRSI, because only reference signals which need to be corrected are changed. Moreover, taking into account these features may prevent erroneous corrections of phase jumps which were induced by eddy currents.

Acknowledgments

The authors thank J.M. Wild for discussions and providing his software. This research is partly funded by the EU: INTERPRET project, IST-1999-10310.

References

[1] RJ Ordidge and ID Cresshull. The Correction of Transient Bo Field Shifts Following the Application of Pulsed Gradients by Phase Correction in the Time Domain, Journal of Magnetic Resonance 1986; **69**: 151-155.

[2] GA Morris. Compensation of Instrumental Imperfections by Deconvolution Using an Internal Reference Signal, Journal of Magnetic Resonance 1988; **80**: 547-552..

[3] JR Roebuck, DO Hearshen, M O'Donnell and T Raidy. Correction of Phase Effects Produced by Eddy Currents in Solvent Suppressed 1H-CSI, Magnetic Resonance in Medicine 1993; **30**: 277-282.

[4] U Klose. In Vivo Proton Spectroscopy in Presence of Eddy Currents, Magnetic Resonance in Medicine 1990; **14**: 26-30.

[5] KR Metz, MM Lam, AG Webb. Reference deconvolution: A Simple and Effective Method for Resolution Enhancement in Nuclear Magnetic Resonance spectroscopy, Concepts in Magnetic Resonance 2000; **12(1)**: 21-42.

[6] HJA in 't Zandt, M van der Graaf and A Heerschap. Common Processing of In Vivo MR Spectra, NMR in Biomedicine 2001; **14(4)**: 224-232.

[7] JM Wild. Artifacts Introduced by Zero Order Phase Correction in Proton NMR Spectroscopy and a Method of Elimination by Phase Filtering, Journal of Magnetic Resonance 1999; **137**: 430-436.

[8] AA Maudsley, Z Wu, DJ Meyerhoff and MW Weiner. Automated Processing for Proton Spectroscopic Imaging Using Water Reference Deconvolution, Magnetic Resonance in Medicine 1994; **31**: 589-595.

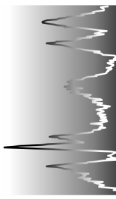
[9] H Barjat, GA Morris, AG Swanson, S Smart and SCR Williams. Reference Deconvolution Using Multiplet Reference Signals, Journal of Magnetic Resonance 1995; Series A **116**: 206-214.

[10] Z Wang, L Bolinger, VH Subramanian and JS Leigh. Errors of Fourier Chemical-Shift Imaging and Their Corrections, Journal of Magnetic Resonance 1991; **92**: 64-72.

[11] L Vanhamme, RD Fierro, S Van Huffel and R de Beer. Fast Removal of Residual Water in Proton Spectra, Journal of Magnetic Resonance 1998; **132(2)**: 197-203.

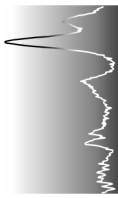
[12] SW Provencher. Estimation of Metabolite Concentrations from Localized in vivo Proton NMR spectra, Magnetic Resonance in Medicine 1993; **30**: 672-679.

[13] L Vanhamme, T Sundin, P Van Hecke and S Van Huffel. MR Spectroscopic Quantitation: A Review of Time-Domain Methods, NMR in Biomedicine 2001; **14**: 233-246.



Chapter 4

Mixture modelling of medical magnetic resonance data‡



Ron Wehrens, Arjan W. Simonetti and Lutgarde M. C. Buydens

Laboratory for Analytical Chemistry, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands

‡ Published in Journal of Chemometrics 2002; **16**, p 1 - 10

Abstract

In clinical decision making, (semi-)automatic unsupervised classification of data for diagnostic purposes is becoming more and more important. The paper describes the application of mixture modelling, a clustering where multivariate gaussians are used to describe clusters in the data, to *in vivo* nuclear magnetic resonance data of patients with brain tumors. Images as well as localized spectra are analyzed. The method is able to automatically generate meaningful classifications. Moreover, the results of clustering both the image and spectral data are in close agreement.

Keywords: model-based clustering, image segmentation, MRI, ^{1}H-MRSI, brain tumor, classification

Introduction

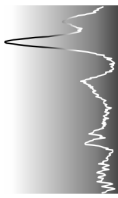
Nuclear magnetic resonance imaging (MRI) has become one of the most important non-invasive diagnostic aids in clinical decision making, mostly because of the good visibility of soft tissue structures. In the diagnosis and treatment of brain tumors, this is of prime importance since it enables the expert to assess location and size of brain tumors. Single images, however, do not always show the desired information and therefore multiple, complementary images are routinely recorded. Although increasing the amount of information, this may hamper diagnosis since radiologists have to investigate multiple images. Therefore, (semi-)automatic methods extracting the relevant features of the images can be of great benefit. Several methods for distinguishing structures in the brain (segmentation) and labeling of these structures have been described in literature[1-3].

A second and potentially even more powerful use of NMR is to record spectra from specific volumes in the brain (usually called proton-magnetic resonance spectroscopic imaging, ^{1}H-MRSI). These volumes are cubes of typically 2 cm^3, called voxels[4]. The spectra, containing information about specific ratios of brain metabolites, can be used to accurately determine the type of tissue in the voxel. Again, automatic or semi-automatic processing of the data can be of tremendous importance. In this paper, an automatic method for the unsupervised classification of both MRI and ^{1}H-MRSI data is described using mixture modelling. The results of the clustering of both types of data show good agreement.

Mixture modelling is an approach to clustering where the data are described as mixtures of distributions, usually multivariate normal distributions[5, 6]. Each gaussian can be considered as one cluster, or a cluster with a non-normal shape can be described by two or more gaussians. Several advantages to model-based clustering over other, more common forms of clustering can be identified:

The clustering has a statistical basis, which allows for inference. It is, e.g., possible to derive uncertainty estimates for individual classifications, as well as for the clustering as a whole.

Several criteria can be used to assess the optimal number of clusters, a direct consequence of the statistical model used to describe the data. This is a large advantage compared to, e.g., hierarchical clustering methods, where a cut-off value must be chosen by the user. In most cases, no clear criteria exist for such a choice.



The clustering method can be selected according to the same criteria used for the choice of the number of clusters. As is the case in hierarchical clustering, several closely related clustering methods exist, and the one that fits the data best can be distinguished in an objective way.

Noisy objects can be explicitly incorporated in the clustering procedure; these objects are then treated as one separate, widely spread cluster polluting the neat grouping of the other objects. In the current application, this feature is not used.

Visualization of the cluster shapes is possible in the space of the original variables. In some cases, this allows for an easier interpretation of the results; e.g., in the current application, neurologists can use domain knowledge to label the different clusters.

Mixture modelling has been applied to a wide variety of clustering problems (for an overview, see[5]), including image segmentation of MRI brain images of healthy volunteers[7]. In this paper, mixture modelling is applied to two types of data sets obtained from five patients with a brain tumor. The first type consists of MRI images of the brains of the patients. The second type of data consists of levels of a small set of metabolites, quantitated in ^{1}H-MRSI data. Although the tumors in many cases can be recognized from the raw data, either the MRI images, or the spectra recorded in the separate voxels, an automatic unsupervised segmentation is very useful to the physician. The individual clusters can be investigated further, e.g. by labeling them as healthy

tissue (e.g. gray matter, white matter), or unhealthy tissue (e.g. tumor, necrotic tissue), or labeling according to the type of tumor. At this stage, supervised methods or expert knowledge may be used. The image segmentation provides an objective definition of the size and location of the clusters, taking into account all data simultaneously. The latter property is important because MRI images can be misleading, especially when only one or two images at the same time are considered: sometimes a tumor seems to be much larger than it actually is, and sometimes it does not show up at all. It is imperative that all data are taken into account, in order to minimize the number of false negatives and false positives.

Mixture modelling is applied to the MRI images of the patients and the levels of the five metabolites. It is shown that tumors can be located in the MRI images as well as in the quantified-metabolites case without human interference. The results for the two types of data show very good agreement. The clustering method is flexible enough to allow for easy assessment of other choices of clustering method or number of clusters.

In the next section, we will briefly review mixture modelling in the context of cluster analysis. Next, the data are presented, followed by the results of clustering both the MRI images and the voxels with the metabolite levels. Although no "true" values for the patients are known (these can only be obtained by an autopsy), the agreement between the clusterings of the two types of data sets indicates that tumors can be located in this way.

Mixture modelling and clustering

Describing a multivariate data set by a mixture of normal distributions is analogous to curve fitting in one dimension, where peaks in, e.g., a chromatogram can (ideally) be described by one or several gauss functions. The sum of these gaussians should fit the original spectrum as close as possible. In mixture modelling, it is the density of the points in multivariate space that should be described by the mixture of normal distributions. This is visualized in figure 4.1.

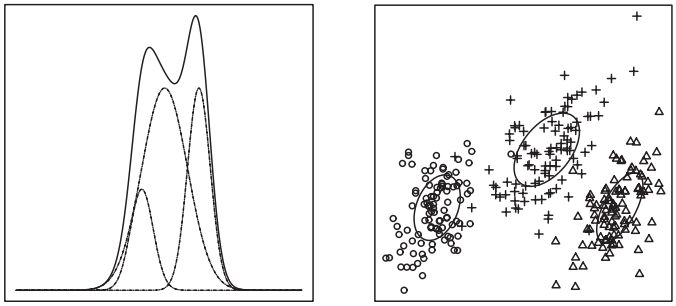


Figure 4.1. The analogy between curve fitting and mixture modelling *In the left panel, a curve (e.g., a chromatogram) is modeled by the sum of several gaussians. In the right plot, the density of points in two-dimensional space is modeled by a mixture of gaussians (indicated by ellipses).*

The $\mathbf{x}$ parameters of these gaussians, means, covariance matrices and proportions, must be estimated from the data. If the class of each object is known, maximum likelihood can be used to obtain these parameters. The density at point $\mathbf{x_i}$ of a multivariate normal distribution, characterized by the mean μ and covariance matrix $\mathbf{\Sigma}$ is given by:

$$\phi(\mathbf{x}_i|\mu,\mathbf{\Sigma}) = (2\pi)^{-p/2}|\mathbf{\Sigma}|^{-1/2}\mathrm{e}^{-(\mathbf{x}_i-\mu)^{\tau}\mathbf{\Sigma}^{-1}(\mathbf{x}_i-\mu)/2} \qquad 1$$

The likelihood $L(\theta|\mathbf{x})$ that all objects are coming from a specific distribution, characterized by $\theta = (\mu, \mathbf{\Sigma})$, is the product of the densities:

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} \phi(\mathbf{x}_i|\theta) \qquad 2$$

The estimates for the distribution describing the data optimally are those that maximize the likelihood $L(\theta|\mathbf{x})$. Many methods can be used for this. Often, the log-likelihood $l(\theta|\mathbf{x}) = \log L(\theta|\mathbf{x})$ is maximized instead of the likelihood.

In the case of mixture modelling, objects may belong to one of several distributions, or clusters:

$$\phi_k(\mathbf{x}_i|\mu_k,\mathbf{\Sigma}_k) = (2\pi)^{-p/2}|\mathbf{\Sigma}_k|^{-1/2}\mathrm{e}^{-(\mathbf{x}_i-\mu_k)^{\tau}\mathbf{\Sigma}_k^{-1}(\mathbf{x}_i-\mu_k)/2} \qquad 3$$

where ϕ_k is the density of the data from cluster k, and μ_k and $\mathbf{\Sigma}_k$ are the mean vector and covariance matrix, respectively, of cluster k. In the case of G clusters, the likelihood is given by:

$$L(\tau, \mu, \mathbf{\Sigma}|\mathbf{x}) = \prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k \phi_k(\mathbf{x}_i|\mu_k, \mathbf{\Sigma}_k) \qquad 4$$

where τ_k is the fraction of objects in cluster *k*.

The problem now is that the classes of the objects $\mathbf{x_i}$ are unknown. The Expectation-Maximization (EM) algorithm by Dempster[8, 9] iteratively solves this problem. The first step in the EM algorithm for mixture models is the calculation of z_{ik} , the conditional probability that object *i* belongs to class *k*, given an initial guess for the parameters θ (= τ, μ, $\mathbf{\Sigma}$):

$$z_{ik} = \phi_k(\mathbf{x}_i|\theta_k) / \sum_{j=1}^{K} \phi_j(\mathbf{x}_i|\theta_j)$$

This is done for all objects and all classes. This is the expectation step; the second step is the maximization step, in which the parameters θ for the mixture model are estimated. The z_{ik} are used in this estimation and therefore these parameters (graphically expressed as the ellipses in figure 4.1) may differ from the initial estimates. The E- and M-steps alternate until convergence. Eventually, objects can be classified to the cluster with the highest z_{ik} for that object: $\max_k (z_{ik})$.

Usually, one does not start the EM algorithm with initial values for θ, but with an initial partitioning and the M-step. Since the choice for an initial partitioning can influence the eventual classification significantly[5], most applications start from a number of different starting points and use the one leading to the best clustering, or use another clustering method to obtain the initial classification. In this case, fast hierarchical methods[10] are used (see software section) for the initialization.

Several restrictions to the shape of the clusters may be imposed. It is, e.g., possible to require that all clusters have the same shape, the same volume, the same orientation or combinations of these[7, 11]. Also completely unrestricted models are possible. Each set of constraints corresponds to a different clustering criterion: if the clusters are restricted to be spherical and identical in size, the criterion that is used equals the criterion in Ward's clustering and *k*-means clustering[6, 12]. Several constraints for which algorithms have been published are gathered in table 4.1; further information on mixture models with restricted covariance matrices may be found in e.g. references[5, 6, 12].

Table 4.1 ***Covariance matrix constraints encountered in mixture modelling. Adapted from reference*** [6]

	distribution	cluster volume	cluster shape	cluster orientation
EI	spherical	equal	equal	-
VI	spherical	variable	equal	-
EEE	ellipsoidal	equal	equal	equal
VVV	ellipsoidal	variable	variable	variable
EEV	ellipsoidal	equal	equal	variable
VEV	ellipsoidal	variable	equal	variable

Although the computations involved can take quite some time, they are quite feasible for limited numbers of objects and variables. If the number of objects is large, one can perform a clustering on a smaller subset and classify the remainder, as already shown by Banfield and Raftery[7]. The subset is used to set up the mixture model by identifying the parameters θ. One application of an E-step then leads to a clustering for all pixels in the image[7, 13]. This strategy does bring the risk that small clusters are missed; however, in the current application the stress is less on the detection of small brain tumors than on the exact localization of different tissue types such as tumor and necrotic tissue. If the number of variables is very large, one can use variable reduction techniques such as PCA. However, in some cases it has been shown that such a PCA transformation can destroy cluster structure[14, 15]. Furthermore, the clusters defined in PC space may not be easily interpretable in the space of the original variables.

Assessing the quality of the clustering

Because the data are described by a statistical model rather than a heuristic procedure, it is possible to choose the optimal number of clusters and the best clustering model. The likelihood of the classification is a first indication, but fails to incorporate the complexity of the model; more complex models will find it easier to fit the data well. Several measures that correct for this are available, of which Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) are the most well-known. The AIC[16] is given by

$$\mathrm{AIC} = -2\log L + 2n_\mathrm{p}$$

where L is the likelihood and n_p the number of parameters in the model (here τ, μ and $\mathbf{\Sigma}$); the model that minimizes the AIC value is picked. The AIC tends to overestimate the number of clusters[5], but it is still often used in practice. The BIC[17] is given by the following equation:

$$\mathrm{BIC} = 2\log L - n_\mathrm{p}\log n$$

where n is the number of objects*. In comparison to the AIC, the BIC will select models with fewer parameters (at least for cases where n>8). Although certain regularity conditions used in the derivation of the measure in practice will not be satisfied, this measure has appeared to work very well. Other methods, including parametric bootstrap methods and cross-validation have been proposed as well (see reference 5 for an overview). In the MCLUST software[13] employed in this paper (see below), the BIC value is used to choose the optimal model and the optimal number of clusters. It should be noted that 'optimal model' and 'optimal number of clusters' are used here in the sense of 'best describing the data'; whether this is also optimal in terms of interpretation of the clustering should be assessed afterwards.

One further consequence of the statistical model behind the clustering is that we can assess the level of confidence in individual classifications. A natural measure of uncertainty for the classification of object i is given by $u_\mathrm{i} = 1 - \max_\mathrm{k}(z_\mathrm{ik})$[5, 6]. Maximal certainty is obtained when $\max_\mathrm{k}(z_\mathrm{ik}) = 1$; because $\mathbf{\Sigma}_\mathrm{k}\, z_\mathrm{ik} = 1$, all other classes have a posterior probability of zero. A measure of the overall uncertainty of the classification is then given by the mean:

* The BIC value is more often used in its negative form, analogous to the AIC. Here, we will adhere to the original definition, as is done in MCLUST.

$$\overline{u} = \frac{1}{n}\sum_{i}^{n} u_i \qquad 5$$

This measure is obviously biased, since the same data are used to set up the model and to estimate the uncertainty. This can be (at least partially) be corrected for, e.g. by bootstrap methods[18]. However, it does give a useful indication on which objects are clustered with relatively large confidence, and which are more uncertain.

Software

Calculations in this paper have been performed by the MCLUST package for model-based clustering, by Fraley and Raftery[13]. Also, figures 4.3 and 4.4 were created with the MCLUST software. The original program (http://www.stat.washington.edu/fraley/mclust/soft.shtml) is an add-on package for S-plus (http://www.insightful.com); here we have used the implementation for R[19]. R, which was used for the other calculations and visualization procedures in this paper, and MCLUST for R, can be downloaded from the R home page (http://www.r-project.org). A more complete account of currently available software for mixture modelling is given in reference 5.

Data

Data for five patients (hereafter denoted with D, G, K, M, and W) with an oligodendroglioma tumor in the brain will be used. The tumors have been pathologically identified by a biopsy, but the exact location and size have not been determined. The first type of data consists of MRI images of four varieties: T_1- and T_2-weighted images, a proton density image and a gadolinium-enhanced (Gd-DTPA) image. In the T_1- and T_2-weighted images the contrast is obtained by differences in T_1- and T_2-relaxation rates, respectively, of the water protons in the brain. In the proton density image, it is the distribution of protons in the brain that is important. The Gd-DTPA image is obtained by injecting a contrast fluid in the blood of the patient. This influences the relaxation of protons, and therefore regions are highlighted containing large blood vessels or regions where the blood-brain barrier is disrupted due to proliferation of malignant tissue.

To avoid misclassification due to non-important regions in the images, the skull and fat tissue surrounding the brain have been removed[20]. Images of each patient are aligned to correct for patient movement during the experiments. Intensities are range-scaled so that the largest intensity is 256 and the smallest (in the brain) is 1; pixels outside the brain are not taken into account in the following analysis. The number of pixels included in the analysis for each of the patients is given in table 4.2.

Table 4.2 Properties of the MRI (top row) and quantified metabolite (bottom rows) data sets for the five patients.

	patient D	**patient G**	**patient K**	**patient M**	**patient W**
no. of pixels	23557	20907	13195	22987	27684
grid	6 x 7	6 x 6	6 x 6	7 x 8	6 x 6
no. of foxels	42	36	33	56	36

For patient K, three corner foxels are not taken into account because they are partially located outside the brain region

From all patients ^{1}H-MRSI data are available, acquired both with and without water suppression using a 16 X 16 2D STEAM ^{1}H-MRSI sequence (Tr = 2500 ms, TE = 20 ms, slice thickness 15 mm, FOV 200 mm, BW = 1000 Hz, NS = 1024 pts). The STEAM box was positioned in the brain and contained, depending on the patient, between 33 and 56 voxels, each with a separate spectrum. The exact number of voxels per patient is given in table 4.2. All spectra are preprocessed in the same automatic way: Eddy current correction[21] to correct for Eddy currents and zero order phase, HSVD filtering[22] to remove the residual water signal between 4.3 and 5.0 ppm, and baseline correction by exponential filtering in the time domain of broad resonances. First order phasing was not necessary because the signal was sampled from the top of the echo. Prior to Fourier Transform, the spectra were zero-filled to 2048 points.

Examples of spectra from normal and tumor areas in the brain of patient M are depicted in figure 4.2. The low signal-to-noise ratio is characteristic for this kind of spectra. In some cases, especially at the outer voxels, effects from outside the grid are noticeable, so that the spectra from the inner voxels are usually of a higher quality.

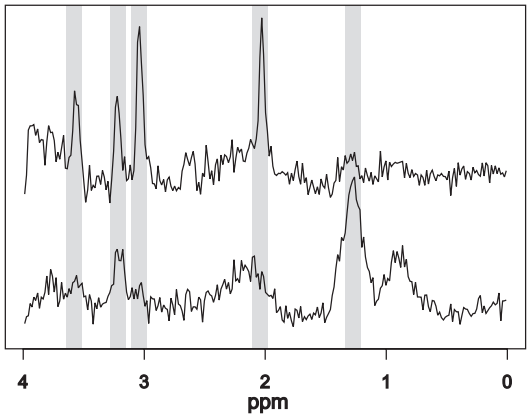


Figure 4.2. *Typical spectra, obtained from patient M, from normal brain tissue (top spectrum, location 9, 5 in figure 4.6) and tumor tissue (bottom spectrum, location 7, 7). The grey bands indicate the integration regions for the five compounds studied.*

A recent overview of NMR data for metabolites that are important for the diagnosis of brain tumors can, e.g., be found in[23]. From the corrected spectra, the levels of five metabolites are determined by integrating the intensities within a window of 17 points (0.13 ppm) at the expected chemical shift. These windows have been indicated with grey bands in figure 4.2. The five components, myo-inositol (at 3.56 ppm), choline (3.20 ppm), creatine (3.02 ppm), N-acetyl-aspartate (NAA, 2.02 ppm), and lactate (1.30 ppm), have been chosen because of their diagnostic importance and the ease of quantification. NAA is hardly ever found outside the nervous system and is considered to be a neural marker. Its levels are shown to decrease in areas with neuronal loss. The choline resonance arises from a group of compounds which are involved in membrane synthesis and degradation. Both increases and decreases in choline levels have been reported in pathological conditions. The creatine signal consists of compounds involved in energy metabolism via the creatine kinase reactions. It is thought to be constant in different brain regions and pathological processes, but can decrease when general cell loss occurs. Lactate is normally below detection limit, but is found to increase in brain tumors and other conditions. Finally, the function of myo-inositol is largely unknown, but it appears to be altered in many pathological conditions. Summarizing: tumor spectra are often characterized by high lactate values and large choline/ creatine ratios[23].

This quantitation procedure is simple and robust, and has the advantage that it will never return completely incorrect results. More elaborate quantitation algorithms, e.g. fitting a Lorentzian at the required location, will sometimes fit the baseline in the absence of a peak. This then leads to completely wrong values, which will not occur with the method used here. Nevertheless, some variation in the levels of the compounds is unavoidable. It should be noted that the levels found with our procedure are only an approximate indication of true concentrations: only part of the peak width is taken into account and negative values are acceptable. For classification purposes, this appears to be adequate.

Results

MRI data. Because the number of pixels in the MRI images is quite large, the classification of MRI images starts with a random selection of 2000 training pixels for each patient. These will be used to build and evaluate the cluster models; the other pixels will be classified with the optimal model. BIC values are calculated from one up to seven clusters. In all cases there is significant structure in the data, showing from the fact that the best BIC value for only one cluster was always significantly lower than the maximal value. As an example, the BIC plot for patient W is shown in figure 4.3.

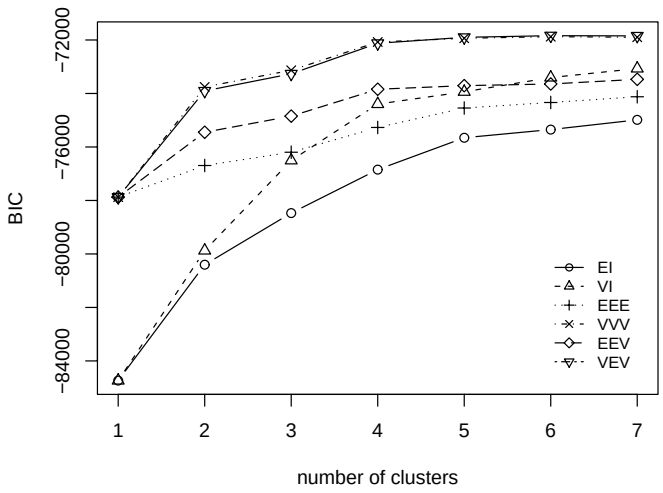


Figure 4.3. BIC plots for the training set of patients W.

The two least constrained models, VVV and VEV, show the highest BIC values. The optimal model is VEV with six clusters. At four clusters a plateau is reached after which the BIC values do not increase very much. A more parsimonious model, therefore, could be the VVV model with four clusters. Similar behavior is observed for the other patients. In table 4.3 the optimal BIC values and the values at which the plateau flattens are gathered. Uncertainty estimates for the classification, calculated by Equation (5), are included as well. With a larger number of clusters, these values are higher, which is to be expected since $\Sigma_k z_{ik} = 1$.

Table 4.3 ***Optimal cluster models identified from the maximal BIC value and optimal models with fewer clusters***

	maximal BIC			start of plateau		
patient	**model**	**BIC**	**uncertainty**	**model**	**BIC**	**uncertainty**
D	VVV 5	-68798	0.086	VVV 4	-69075	0.061
G	VVV 5	-36150	0.096	VVV 4	-36224	0.082
K	VEV 7	-73149	0.158	VVV 3	-73537	0.089
M	VVV 7	-67117	0.111	VVV 5	-67474	0.078
W	VEV 6	-71839	0.107	VVV 4	-72074	0.054

The training set and the definition of clusters in the first two dimensions (the T_1- and T_2-weighted images) of patient W are shown in figure 4.4. The optimal model contains six clusters, indicated by the ellipses. The clustering model is VEV, indicating that all ellipses have the same shape but differ in volume and orientation. The cluster with high T_1 values corresponds to the tumor. Unfortunately, this is not always the case, so for the identification of the clusters other sources of information are necessary.

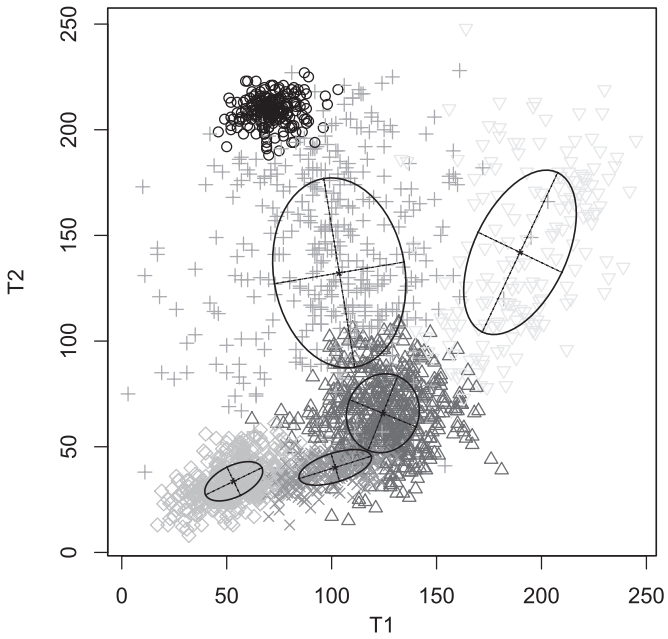


Figure 4.4. Definition of clusters in the space of the first two images. *Each point corresponds to a pixel. The ellipses indicate the six gaussians that describe the data optimally, according to the BIC criterion. The model is VEV, where the shape of all clusters is equal (although it seems to be different, due to the projection in two-dimensional space), and size and orientation are variable.*

The classifications corresponding to the optimal BIC values are depicted in figure 4.5. Using knowledge of the structure of the brain and information on the patients, the separate clusters could be assigned to specific brain tissues. Clusters describing one type of tissue have been plotted in the same color. In all cases, the tumor, indicated in black, is clearly visible. For clarity, grey and white matter are both depicted in white but in the original color-coded pictures they are clearly distinguished. The cerebro-spinal fluid (CSF) is indicated in grey. The tumor pixels consist of elements from a single cluster, except in the case of patient K, where the tumor is described by two gaussians. For all patients, two clusters are used to describe the CSF. If clusterings with fewer clusters are used, tumors and CSF are always described with only one cluster.

The tumor in patient K is the least well defined: tumor pixels are found in almost all regions of the brain of the patient, although the bulk of the tumor is in the lower left region. Patient W shows tumor pixels at several locations as well, but again the bulk of the tumor cluster is in a closely defined region. The three other patients show well-defined tumor regions. Patient M shows two CSF regions within the tumor: the top left region is known to result from a bleeding in that area.

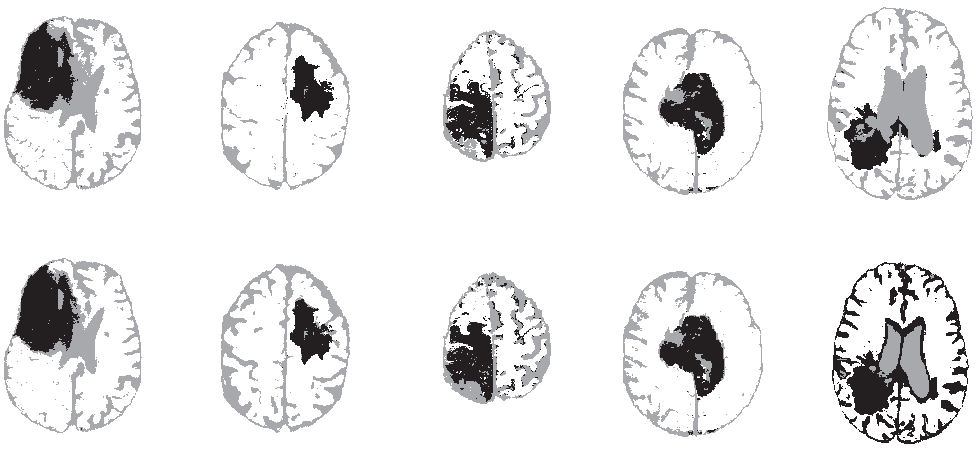


Figure 4.5. Automatic segmentations of MRI images from patients D, G, K, M and W (left to right). *Tumors are indicated in black; cerebro-spinal fluid in grey. Top row: optimal models; bottom row: models with fewer clusters from table 4.3.*

As a measure of agreement between the 'optimal' and more parsimonious models, we calculated the percentage of pixels in the optimal clustering that was classified similarly in the more parsimonious clustering. These percentages, for the tumor and CSF clusters, are given in table 4.4, indicating that the choice of model and number of clusters is not very sensitive, as long as reasonably high BIC values are chosen. The BIC values differ by as much as 300 units, but still the classifications are more or less similar. For all patients, more than 90 percent of all pixels that are classified as tumor in the optimal models are also classified as tumor in the clustering with fewer clusters. The CSF classification for patients K and M is also in very close agreement; in patient D the CSF layer at the outer edge of the skull is somewhat thicker in the optimal BIC clustering, and also shows a bit more detail. The reverse is true for patient G, where the four-cluster solution shows a thicker CSF layer than the (optimal) five-cluster solution. The CSF clustering with four clusters

for patient W deviates significantly from the one with six clusters because the tumor cluster also contains a large part of the CSF, while the CSF in the butterfly-like structure in the centre of the brain is classified separately.

The influence of the choice of the training set can be significant in terms of the model that is chosen and the number of clusters that appears optimal. However, repeated application of the clustering procedure in this case shows only minor differences in the optimal model or the number of clusters; the classification of the pixels changes very little as well. Differences are larger when smaller training sets are used; in experiments where only 500 training pixels were selected, the optimal number of clusters could range from four to seven for different training sets. The major structure, however, was retained in all cases.

Table 4.4 ***Percentage of tumour and CSF pixels respectively of the optimal models that are calssified identically in the models with fewer clusters***

patient	tumour (%)	CSF (%)
D	99.8	80.7
G	93.7	100.0
K	95.9	94.7
M	98.3	94.6
W	100.0	34.1

Metabolite data. Because the number of voxels for which quantitation has been done is much smaller than the number of pixels in the images, all voxels can be taken into account when constructing the clustering models. All models with three to five clusters were considered; more clusters would not be informative, given the small number of voxels, and fewer clusters would probably not discriminate tumor tissue from healthy tissue. The results of the optimal models are shown in table 4.5 and figure 4.6. The tumor clusters are indicated in dark grey, the other clusters in lighter shades. Uncertainties in the classification are indicated as well. In most cases, the region of the tumor is classified with little uncertainty.

Table 4.5 Optimal cluster models for the metabolite data of the five patients

	D	G	K	M	W
optimal model	VI 5	EEE 3	EI 3	EEV 4	EEE 3
no. of clusters per tumour	1	2	1	2	1
no. of tumour voxels	10	12	3	18	4

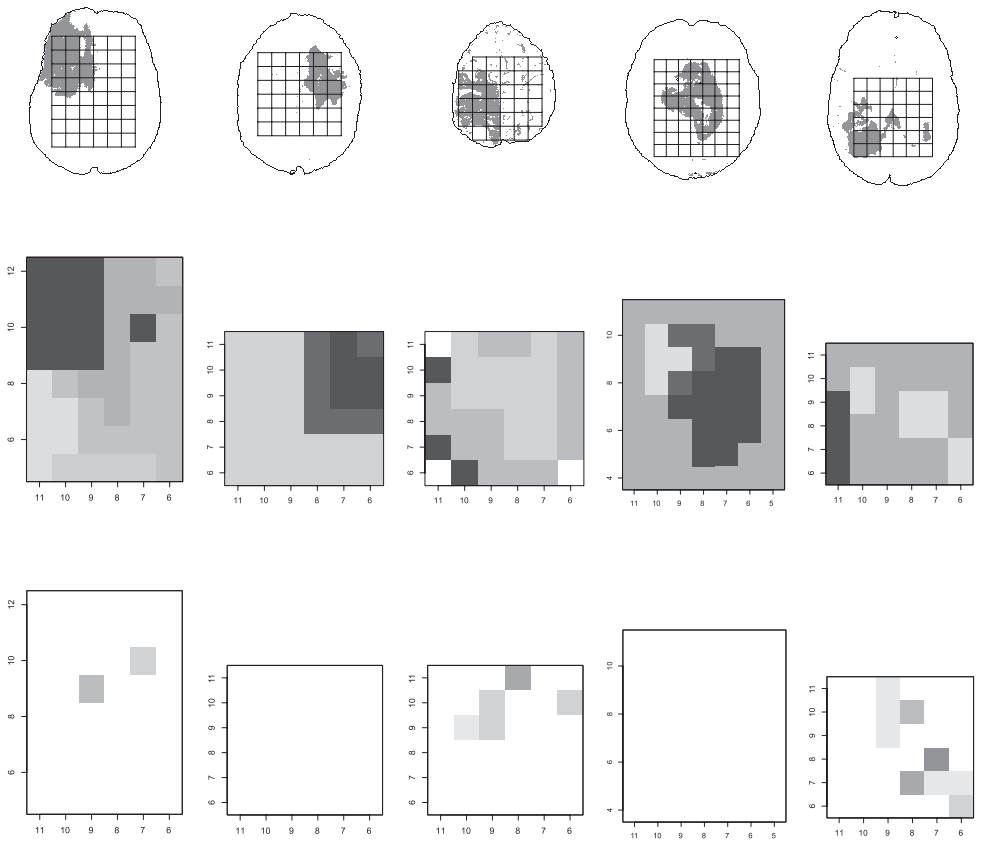


Figure 4.6. Automatic clustering of brain tumors using metabolite levels of patients D, G, K, M and W. *Top row: location of tumor, according to MRI segmentation, and voxels. Middle row: optimal classification with three to five clusters. Three corner voxels of patient K are not taken into account: these are indicated in white. Bottom row: uncertainties in the classification. Dark values indicate larger uncertainties. The largest uncertainties are 0.29, 0.02, 0.45, 0.02 and 0.49 for patients D, G, K, M and W, respectively.*

Identification of each cluster is performed by looking at the mean component levels in the clusters. As an example, the box plot for patient W is shown in figure 4.7. Clearly, the first cluster constitutes the tumor (high lactate values and a large choline/creatine ratio), while the other two clusters constitute normal brain tissue.

Instead of analyzing data from each patient separately, one can also pool the patients and classify all of them at once. This is only possible if there are no major inter-patient differences. In this case, box plots of metabolite levels for these five patients showed that differences were not very big. Indeed, the automatic classification of the pooled patients showed clearly the same tumor regions as the patient-per-patient classification. All models with one to seven clusters were evaluated. The model with the highest BIC value was EEV with six clusters. The classification with this model and the associated uncertainties, are shown in figure 4.8; two tumor clusters were identified. The area of the bleeding in patient M is classified as a completely separate cluster. Again, the regions of the tumors are classified quite confidently. Uncertainties are somewhat higher than in the patient-per-patient classification, which is to be expected because of natural variation between patients. On the other hand, all the patients are now classified with more or less the same uncertainty.

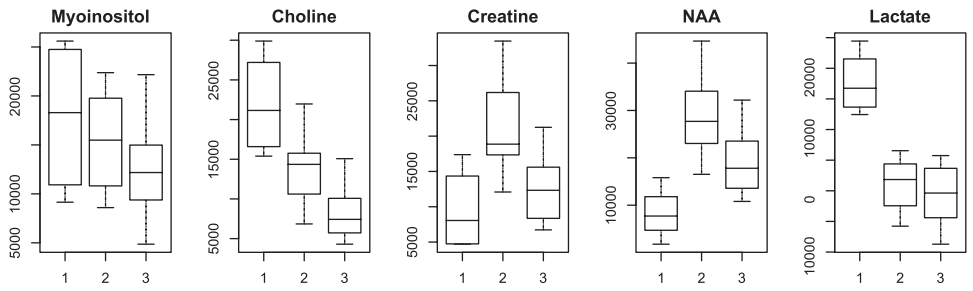


Figure 4.7. Metabolite levels in the clusters of patient W. *The first of the three clusters corresponds with the tumor. The horizontal line in a box depicts the median, the box itself is bounded by the upper and lower quartiles of the data, and the whiskers extend to the most extreme data points not more than 1.5 times the interquartile range from the box.*

Comparison of results. A comparison with 'true' values is impossible, but further insight can be obtained when the clusterings based on the two data types are compared with each other. We concentrate on the tumor clusters.

Although the MRI images provide more detail, they are not always a better indicator of tumor size and location. An example can be seen in the classification of patient M in figures 4.6 and 4.8. In all cases where the voxels are dark, there is a corresponding tumor classification in the MRI image. However, the top part of the MRI tumor cluster is not recognized as a tumor in the metabolite classification. This may be caused by gadolinium leaking into a cavity in the brain, with a high intensity in the Gd-enhanced image as a result. The classification of the tumor in the image may therefore cover too large a region. The four light-shaded voxels in the top left of the tumor in figure 4.6 correspond to an area in which a bleeding has occurred; this is also visible in the MRI classification. Similar findings are visible for patients D and W: not all voxels covering the tumor cluster in the MRI image are classified as tumor tissue.

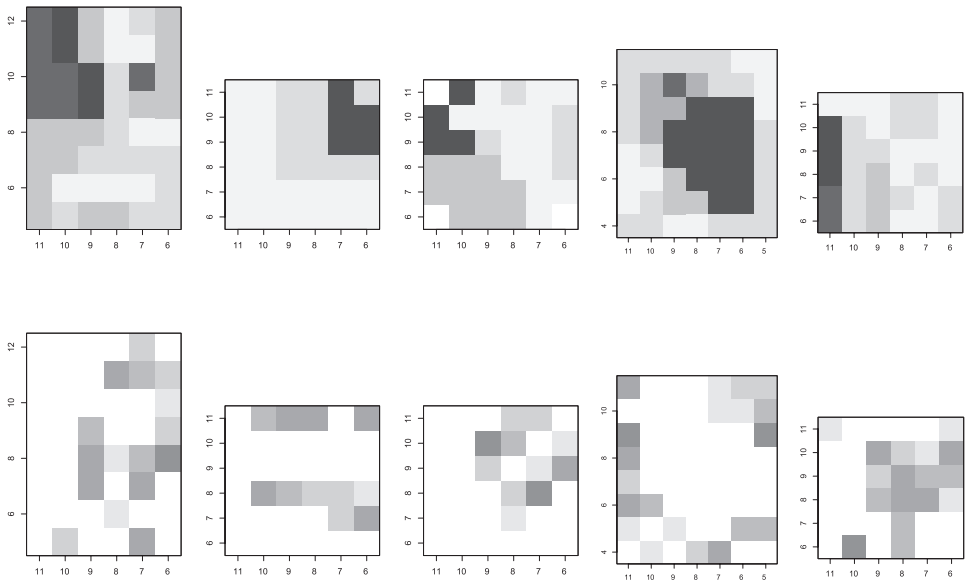


Figure 4.8. Classification based on pooled patient data. *Top: optimal model with six clusters (EEV). Dark voxels correspond with tumor tissue. For patient K, three of the corner voxels are not classified because they are outside the brain (depicted in white). Bottom: uncertainty in the clustering. Dark voxels correspond with higher uncertainties. The largest uncertainties for the five patients are 0.33, 0.24, 0.30, 0.35 and 0.44, respectively.*

The patient-per-patient metabolites classification of patient K shows a large difference with the MRI classification. Whereas two of the three tumor voxels coincide with the tumor cluster in the MRI image, the third voxel contains mostly CSF, according to the MRI. Moreover, the majority of the tumor pixels do not have features characteristic for tumor in their spectrum. The pooled-patient classification is in better agreement with the classification of the MRI data. This may be the result of a better definition of the tumor cluster, taking into account metabolite levels of the other patients. A possible reason for the relatively bad agreement may be the fact that the grid is acquired high in the brain and therefore lies close to the skull. This can cause artifacts arising from fatty acids in tissue surrounding the brain.

Patient G shows the nearly perfect agreement between the MRI classification and the metabolites classification: in the patient-per-patient classification, voxels completely overlapping with the tumor cluster in the MRI images form one tumor cluster. Voxels partially overlapping form a second cluster, still with tumor characteristics, so the boundary of the tumor forms a separate class. The agreement with the pooled-patient data is also very good; this goes for all patients with the exception of patient K.

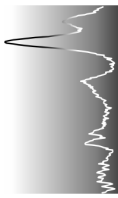
With the MRI data it was not possible to pool the patient data. The absorption levels differed too much between the patients, resulting in e.g. the tumor of patient D being classified in the same cluster as the CSF of patient M. Nevertheless, the overall structures were very similar to the results of the patient-per-patient classification, and a suitable scaling of intensities in the images could very well lead to a correct overall clustering. This falls outside the scope of this paper.

Conclusion

Mixture modelling is a powerful and flexible technique for clustering in cases where the number of dimensions is not too large. In such cases, direct interpretation of the cluster shapes in the space of the original variables is possible. Large numbers of objects can be easily tackled by specifying training sets to set up the model, which in turn can be used to classify the other objects. This makes the method very suitable for use in such fields as data mining[24, 25] or image analysis[26]. Of course, the training set should be representative for the other data; more specifically, all clusters should be well represented.

The problem of selecting the optimal clustering method and the optimal number of clusters is usually tackled by a combination of common sense, domain knowledge and trial and error. This, unfortunately, adds a level of subjectivity which can make it difficult to reproduce or interpret results. In this paper, the use of BIC values, which are used in MCLUST as indicators for clustering method and number of clusters, is assessed. Ultimately, there is no guaranteed way to assess the quality of a partitioning other than post-hoc validation. Several different partitionings should be looked at before deciding which one is best. Nevertheless, all classifications in this paper using the highest BIC values were very good.

In this paper, we have not taken into account the spatial relations between objects that are clustered (pixels and voxels). If objects within one cluster are also located close to each other in, e.g., an image, this leads to a higher confidence that the partitioning is meaningful. Thus, domain knowledge can be used to obtain at least a partial validation of the clustering. Another approach is to include the distance of pixels in the image in the clustering, so that one increases the chance that nearby pixels are classified in the same cluster. How to do that in the context of model-based clustering is as yet unclear.



Although no "true" classifications were available, the use of two different types of data of the same patients confirmed that the clusterings described the structure of the data well. The identity of clusters found in the MRI images could be established by aligning the images with the tumors identified in the coarser grid of voxels. In this way, the advantages of both methods are combined: the large spatial resolution of MRI imaging with the high information content of the ^{1}H-MRSI data. Mixture modelling is an attractive way to obtain such clusterings.

Acknowledgements

J. Noordam and an anonymous referee are acknowledged for constructive comments on an earlier version of the manuscript. Partial funding for the research in this paper has been obtained from the EC-funded INTERPRET project (IST-1999-10310).

References

[1] L.P. Clarke, R.P. Veldhuizen, M.A. Camacho, J.J. Heine, M. Vaidyanathan, L.O. Hall, R.W. Thatcher, and M. Silbiger. MRI segmentation: methods and applications. Magn. Reson. Imaging 1995; **13(3)**: 343-368.

[2] W.E. Reddick, J.O. Glass, E.N. Cook, T.D. Elkin, and R.J. Deaton. Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks. IEEE Trans. Med. Imaging 1997; **16**: 911-918.

[3] H. Witjes, A.W. Simonetti, and L.M.C. Buydens. Better brain imaging with chemometrics. Anal. Chem. 2001; **73(19)**: 548A-556A.

[4] S.M. Noworolski, S.J. Nelson, R.G. Henry, M.R. Day, L.L. Wald, J. Star-Lack, and D.B. Vigneron. High spatial resolution 1H-MRSI and segmented MRI of cortical gray matter and subcortical white matter in three regions of the human brain. Magn. Reson. Medicine 1999; **41**: 21-29.

[5] G.J. McLachlan and D. Peel. Finite mixture models. John Wiley & Sons, New York, 2000.

[6] C. Fraley and A.E. Raftery. How many clusters? which clustering method? - answers via model-based cluster analysis. Computer J. 1998; **41**: 578-588.

[7] J.D. Banfield and A.E. Raftery. Model-based gaussian and non-gaussian clustering. Biometrics 1993; **49**: 803-821.

[8] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. J. Royal Statist. Soc. B 1977; **39(1)**: 1-38.

[9] G.J. McLachlan and T. Krishnan. The EM algorithm and extensions. Wiley, 1997.

[10] C. Fraley. Algorithms for model-based gaussian hierarchical clustering. SIAM J. Sci. Comput., 20:270-281, 1998. Also: Technical Report no. 311, Department of Statistics, University of Washington.

[11] F. Murtagh and A.E. Raftery. Fitting straight lines to point patterns. Patt. Recogn. 1984; **17**: 479-483.

[12] G. Celeux and G. Govaert. Gaussian parsimonious clustering models. Patt. Recogn. 1995; **28(5)**: 781-793.

[13] C. Fraley and A.E. Raftery. MCLUST: Software for model-based cluster analysis. J. Classif. 1999; **16(2)**: 297-306. Also: Technical Report No. 342, Dept. of Statistics, University of Washington.

[14] W.C. Chang. On using principal components before separating a mixture of two multivariate normal distributions. Appl. Statist. 1983; **32(3)**: 267-275.

[15] C. Ricotta, G.C. Avena, and F. Volpe. The influence of principal component analysis on the spatial structure of a multispectral dataset. Int. J. Remote Sens. 1999; **20(17)**: 3367- 3376.

[16] H. Akaike. A new look at the statistical model identification. IEEE Trans. Automatic Control 1974; **19**: 716-723.

[17] G. Schwarz. Estimating the dimension of a model. Ann. Statist. 1978; **6**: 461-464.

[18] K.E. Basford and G.J. McLachlan. Estimation of allocation rates in a cluster analysis context. J. Am. Stat. Assoc. 1985; **80**: 286-293.

[19] R. Ihaka and R. Gentleman. R: A language for data analysis and graphics. J. Comput. Graphic. Statist. 1996; **5(3)**: 299-314.

[20] Z. Liang, D. Wang, J. Ye, and D. Harrington. Development of automatic techniques for segmentation of brain tissues from multispectral images. IEEE Nucl. Sci. Symp. 1995; **3**: 1453-1456.

[21] U. Klose. In-vivo proton spectroscopy in presence of Eddy currents. Magn. Reson. Medicine 1990; **14**: 26-30.

[22] L. Vanhamme, R.D. Fierro, S. van Huffel, and R. de Beer. Fast removal of residual water in proton spectra. J. Magn. Reson. 1998; **132**: 197-203.

[23] V. Govindaraju, K. Young, and A.A. Maudsley. Proton NMR chemical shifts and coupling constants for brain metabolites. NMR in Biomedicine 2000; **13**: 129-153.

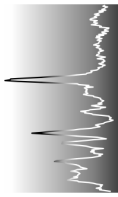
[24] P. Smyth. Data mining: data analysis on a grand scale? Technical Report UCI-ICS 00-20, University of California, Irvine, July 2000.

[25] F. Murtagh. Clustering of massive data sets. In J. Abello, P.M. Pardalos, and M.G.C. Reisende, editors, Handbook of massive data sets. Kluwer, in press.

[26] D.A. Landgrebe. Information extraction principles and methods for multispectral and hyperspectral image data. In C.H. Chen, editor, Information Processing for Remote Sensing, chapter 1. World Scientific Publishing Co., Inc, River Edge, NJ, USA, 2000.

Chapter 5

A new chemometric approach for brain tumor classification using magnetic resonance imaging and spectroscopy‡



Arjan W. Simonetti[♣]*, Willem J. Melssen*[♣]*, Marinette van der Graaf*[♦]*, Geert J. Postma*[♣]*, Arend Heerschap*[♦] *and Lutgarde M. C. Buydens*[♣]

[♣]**Laboratory for Analytical Chemistry, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands**

[♦]**Department of Radiology, University Medical Center Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands**

‡ Published in Analytical Chemstry 2003; **75**(20), p 5352 - 5361

Abstract

A new classification approach was developed to improve the non-invasive diagnosis of brain tumors. Within this approach information is extracted from magnetic resonance imaging and spectroscopy data, from which the relative location and distribution of selected tumor classes in feature space can be calculated. This relative location and distribution is used to select the best information extraction procedure, to identify overlapping tumor classes and to calculate probabilities of class membership. These probabilities are very important, since they provide information about the reliability of classification and might provide information about the heterogeneity of the tissue. Classification boundaries were calculated by setting thresholds for each investigated tumor class, which enabled the classification of new objects. Results on histopathologically determined tumors are excellent, demonstrated by spatial maps showing a high probability for the correctly identified tumor class and, moreover low probabilities for other tumor classes.

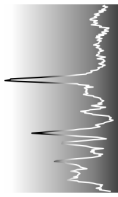
Introduction

Magnetic Resonance Imaging (MRI) has become one of the most important non-invasive aids in clinical diagnosis of brain tumors, because it enables the radiologist to assess the anatomical location and size of tumor tissue. Normally, multiple complementary images are recorded, as the acquisition of only one image may not show the desired information. Another use of clinical MR is to perform spatially resolved Magnetic Resonance Spectroscopy (MRS). In contrast to MRI, which has a high spatial resolution, this method is performed at low spatial resolution, but it involves the acquisition of metabolic information. The method allows the use of several MR sensitive nuclei, but often the proton nucleus is used as it is the most sensitive one. Proton MRS at multiple spatially resolved locations is referred to as Proton Magnetic Resonance Spectroscopic Imaging (^{1}H-MRSI)[1].

When applying ^{1}H-MRSI on the human brain, the MR spectra obtained on a clinical 1.5 T MR system commonly arise from volumes (called voxels) of approximately 2-4 cm^3. The spectra contain information from important brain metabolites like glutamate, myo-inositol, choline, creatine, N-acetyl aspartate (NAA), lactate and fatty acids[2]. NAA is considered to be a neuronal marker of which the levels are decreased in brain tumors. The choline resonance arises

from a group of compounds which are involved in membrane synthesis and degradation, which is often increased in tumors. The creatine signal reflects compounds involved in energy metabolism and may be decreased in brain tumors. Lactate is normally below detection limit, but it is often increased in brain tumors. The function of myo-inositol is largely unknown, but its signal appears to be altered in some pathologies. The signals from fatty acids are known to be increased in several brain tumors, especially high grades, and are an indicator for necrosis (cell death). Glutamate is an important neurotransmitter in brain and elevated in oligodendrogliomas [3].

The MRSI spectra could be used collectively with the anatomical MR images to spatially determine the type of tissue with a high reliability. However, in a clinical setting the increased amount of information hampers the diagnosis, since radiologists have to investigate multiple MR images and spectra. This is time consuming and often clinicians are not familiar with spectral information. To deal with this problem, a project called INTERPRET (http://carbon.uab.es/INTERPRET) was funded by the European commission. This project had the goal to develop new methodologies to automatically type tumors in the human brain with the use of MRS and MRI data. The resulting decision support system enables clinicians to utilize MRS information for the diagnosis of brain tumors. In this paper a new approach is presented to explore the information content of the combined MRI and MRSI data, collected from patients with a brain tumor. Image like maps can be generated which show the classification of the tissue and the probability of classification in order to assist the clinician in the interpretation of the data.



The development of statistical models (especially multivariate ones) which are focused on the classification of brain tumors from MRSI data is rather new. Some attempts have been made [4-7], but a dedicated multivariate determination of the location and distribution of classes in feature space has not been performed yet. Both, Preul [4] and Szabo de Edelenyi [6] focus on the quantitation of metabolite regions from the spectroscopic data, followed by linear discriminant analysis. Others have used pattern recognition [8, 9] on single voxel MR spectra to classify [10-13] or cluster [14, 15] multiple tumor classes. The emphasis of data analysis is often put on pre-processing [16, 17], data reduction and the search for appropriate classification algorithms. However, a very important part of a classification approach has not received much attention yet. This deals with the assumption of multivariate normality, the identification of class boundaries for each tumor class, the recognition of overlap between

classes and the probability of classification. Multivariate normality is important to consider if the classification method assumes it, especially when also probabilities of class membership of a certain class are to be given. The identification of class boundaries is important since there are many different types of brain tumors. The usually limited number of investigated tumor classes in a classification procedure should therefore have strict class boundaries. This leaves room for the non investigated classes in feature space, which will (hopefully) be classified as 'unknown' by the classifier. The probability of classification is important, because it provides information on the reliability of classification.

Few papers deal with the combination of MRI and MRSI data (Szabo de Edelenyi [6] uses information from the T_2-weighted image, combined with MRSI). If a patient with a brain tumor is investigated, always MR images with different contrasts are acquired. If also a MRSI measurement is performed, it would be a waste to leave out the MRI information in the statistical models, since the MR images are already available. It is possible to identify morphological different regions in the brain with MRI information alone [18]. If this information is combined with MRSI data, the pathology of the regions might also be identified. Therefore, we combined MRI and MRSI data in all our experiments.

The goal of this paper was three fold. First, to find an appropriate feature selection technique for the MRSI data, while preserving the distances and distribution of the classes in feature space. We compared two techniques, simple quantitation of spectral regions and principal component analysis. The second goal was to find the isolated and overlapping classes of tissue types in feature space and to create plots which showed their distribution. These distribution plots can be used to generate spatial probability maps, which show for each voxel the probability that it belongs to a certain tissue class. The third goal was to develop a classification model, which uses thresholds which are set with use of the distribution plots. These thresholds define class boundaries for each class, enabling the classification of voxels to a specific tissue type or as 'unknown'.

To address the above research issues, a training set was setup using data from volunteers and patients with a brain tumor. The construction of the distribution plots was performed with a cross validation method. The evaluation of the classifier was performed with the leave one out method on the training set. Furthermore, the resulting classification maps of the patients

were visually investigated. Most of the voxels within each classification map of a patient were not selected within the training set. If the classification map of a patient contained homogeneous regions of healthy tissue, cerebro-spinal fluid and tumor tissue, then this demonstrated robustness of the system to a large extent. Next to this, two patients from which no data was selected for the training set were validated.

Experimental section

Data. Data from 18 patients with a brain tumor and 4 volunteers selected from the INTERPRET [19] database were used. Each patient case had passed strict quality control and validation procedures, including consensus histopathological determination of the tumor according to the rules of the World Health Organization. For two of the 18 patients (I-1172 and I-1281) a consensus was not reached. Therefore, data from the tumor region of these patients was not used in the training set.

For each subject, stacked MR images of cross-sections of the whole brain at four contrasts were acquired: T_1- and T_2- weighted images, a proton density image and a gadolinium-enhanced T_1-weighted image (256 x 256, FOV = 200 mm, slice thickness = 5 mm). In the T_1- and T_2-weighted images, the contrast reflects differences in T_1- and T_2-relaxation times of the water protons in the brain, respectively. In the proton density image, the distribution of protons produces the contrast. The contrast enhanced image (GD-image) is obtained after intravenous administration of Gd-DTPA solution. The paramagnetic Gd compound influences the relaxation times of water protons and therefore regions are high-lighted which contain large blood vessels or where the blood-brain barrier is disrupted due to proliferation of malignant tissue. To ensure that image pixels from subsequent images originate from the same spatial location, the images have to be co-aligned. The co-alignment was performed by calculating the maximum correlation between two images while shifting them left-right and up-down. Co-alignment is difficult, but does not need to be perfect in this method, because the high resolution images are combined with the low resolution MRSI eventually.

Besides the acquisition of images, also ^{1}H-MRSI data was acquired from each subject, both with and without water signal suppression using a 2D STEAM ^{1}H-MRSI sequence (16 x 16 x 1024 matrix size, Tr = 2000 or 2500 ms, TE = 20 ms, slice thickness 12.5 or 15 mm, FOV = 200 mm, spectral

width = 1000 Hz, NS = 2 (for the MRSI without water suppression, NS = 1)). When STEAM [20] is applied, a box is selected from which the signal is acquired by the application of three slice-selective radiofrequency pulses in the presence of x-, y-, and z-gradients. Only the MRSI voxels within the STEAM box will contain signal, while signals in voxels from outside the box are suppressed. The STEAM box was positioned totally in the brain, therefore disturbing signals arising from the fat tissue surrounding the skull were avoided. The location of the ^{1}H-MRSI slice was selected and aligned with the Gd-image which showed the largest enhanced tumor area. Thus, the MRSI slice (of at most 15 mm thickness) was centered around a MRI slice of 5 mm. Since the MRI slices above and below the selected MRI slice have a gap of 1.5 mm with the selected slice, they are partially outside the view of the MRSI slice. Therefore, only MR images from one slice location have been used.

Proper pre-processing of the MRSI data is essential for successful classification. All data were processed in the same automatic way. In short: k-space data (16 x 16) was filtered by a hanning filter of 50% using the LUISE software package (Siemens, Erlangen, Germany), followed by zero filling to 32 X 32 and 2D Fourier transformation to obtain time domain signals for each voxel. Then, each voxel within the pre-selected STEAM box was corrected for eddy current effects in the spectra using a method which prevents the occasional occurrence of eddy current correction induced artifacts [21]. This was followed by HLSVD [22] filtering to remove the residual water signal between 4.3 and 5.5 ppm. Next, a simple but efficient baseline correction was applied, which removed broad resonances. This was performed by exponential filtering of the time domain signal of each voxel with a filter width of 5 ms, followed by the subtraction of the residual from the original time domain signal. To correct for first order phase effects (in principle the zero order phase is removed by the eddy current correction), the mean spectrum which was calculated from all spectra within the STEAM box of each patients MRSI dataset was manually optimized. This correction value was then applied to each separate time domain signal of that dataset. Finally, each time domain signal was Fourier transformed and only the real values of the complex spectrum was retained between 0.5 to 4.0 ppm (230 points). To normalize between patients, all spectra of one dataset were divided by the patients mean water signal of all voxels within the STEAM box, as calculated by singular value decomposition [23] from the unsuppressed MRSI data.

Next, the spatial resolution of the MR images and the MRSI grid had to be readjusted, so that the two were in agreement. Therefore, the resolution of the MR images was lowered to the resolution of the MRSI grid by averaging the image pixels within each spectroscopy voxel. The values within each low resolution image were scaled to the same range as the spectral data. After pre-processing, each voxel within the grid was represented by a spectrum of 230 variables (the region between 0.5 to 4.0 ppm) and 1 variable from each MR image. This is schematically represented in figure 5.1a-c.

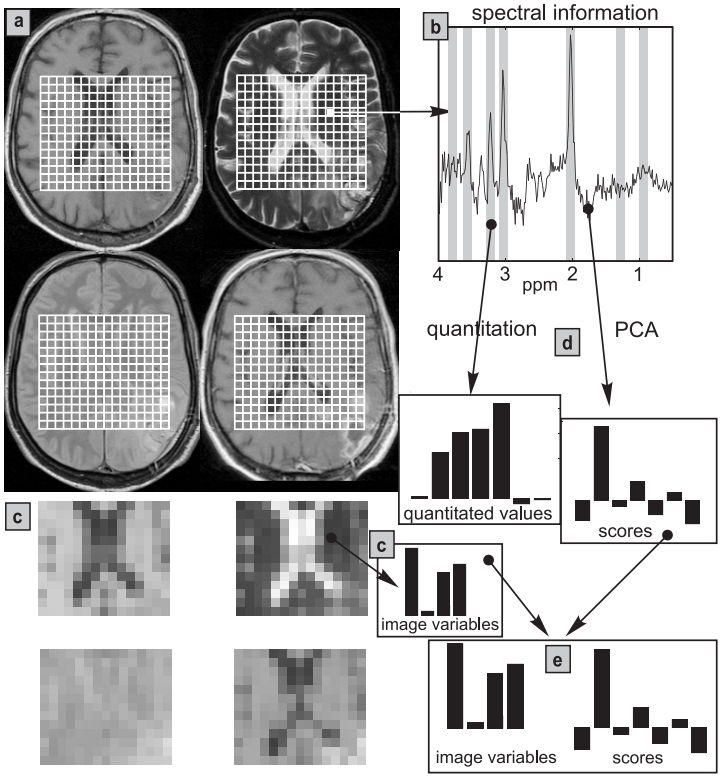


Figure 5.1. Representation of pre-processing and data reduction. *a) The images with the four contrasts (top: T_1- and T_2- weighted image, bottom: proton density and Gd-image) are aligned (the lower right image has been shifted up) and the spectral grid within the STEAM box has been superimposed on the images; b) a spectrum after pre-processing, obtained from the indicated voxel; c) the resolution of the images is reset to the MRSI resolution. Each voxel is represented by four image variables now; d) data reduction is performed either with quantitation of important regions in the spectrum (indicated with gray bands in figure 5.1b) or with PCA of the spectra; e) the image variables from each voxel are combined with the variables obtained from either data reduction technique. In this example the image variables are combined with the variables (scores) obtained by PCA.*

Setup of a training set. Based on the 16 patients for which consensus about the histopathology was reached, four different tumor classes were selected. The first three classes contained glial tumors with different grades, grade II (6 patients), grade III (4 patients) and grade IV (5 patients). A fourth class consisted of a patient (I-1283) with a meningioma (MNG). No classifier for this class was calculated, because the size was to small, but it was used to validate results of the classifier.

To create the training set, a selection of voxels from patients who were in the same tumor class was made using the MR spectra and the four MR images. The voxels were selected from regions which clearly consisted of tissue belonging to the correct class. Boundary voxels were not selected. Also classes for healthy tissue and cerebro-spinal fluid (CSF) were created. The voxels for healthy tissue were selected from the contralateral region of several patients and from healthy volunteers. Data for the CSF was selected from CSF voxels which were not in close contact with the tumor region. The total number of voxels selected from a subject was never more than 30%.

Patient I-1172, for which no consensus was reached, was used for the selection of training set data for healthy tissue and cerebro-spinal fluid. Data from patient I-1281 was not used for the training set at all, but was used as an independent test case for evaluation, just like patient I-1283 (the MNG case). An overview of the training set is given in table 5.1.

Table 5.1. Selected patients, volunteers and number of voxels for each class in the training set.

tumor class	patient number (I-) [a]	no. of voxels selected	total
grade II	1227 1233 1260 1275 1278 1318	07 20 28 12 14 09	90
grade III	1212 1229 1230 1232	17 06 26 4	53
grade IV	1231 1234 1285 1286 1322	12 06 16 17 08	59
MNG	1283	15	15
healthy (volunteer)	1236 1237 1238 1239	29 31 25 37	122
healthy (patient)	1172 1212 1231 1285 1318	13 32 21 39 17	122
CSF	1172 1230 1231 1234 1275 1318	15 09 26 03 09 16	78
no consensus	1281	used as independent test case	
		total	539

[a] *The patient numbers correspond with the INTERPRET [19] database.*

Data reduction methods. After selection of the training set, two different data reduction techniques were used to create two separate training sets with reduced dimensionality, which were validated for their robustness. The following techniques were used:

a) Quantitation of important spectral regions. Quantitation was performed in the following way: from the pre-processed spectra, intensities of seven spectral ranges encompassing the brain metabolites mentioned in the introduction were determined by integrating the intensities within a window of 0.13 ppm at 3.75 ppm (glutamate), 3.56 ppm (myo-inositol), 3.20 ppm (choline), 3.02 ppm (creatine), 2.02 ppm (NAA), 1.33 ppm (lactate), and 0.90 ppm (fatty acids). These seven windows have been indicated with gray bands in figure 5.1b. The quantitation procedure is simple, robust and easily automated. It should be noted that the levels found with our procedure are only an approximate indication of the true levels: only part of the peak width is taken into account and negative values are acceptable in case of a negative baseline. However, for classification purposes this appears to be adequate. The seven spectroscopic

variables (figure 5.1d) and the four image variables (figure 5.1c) were combined to create a training set which contained 11 variables for each voxel.

b) principal component analysis of the spectral data. The second data reduction technique used principal component analysis (PCA) [24]. PCA was performed on the spectra of the training set, which consisted of 524 (MNG excluded) spectra with 230 variables each. PCA finds directions in the data which explain most of the variance. These new axes are linear combinations of the original ones, are orthogonal and ordered with respect to the amount of explained variance. After PCA, the first seven principal components, which explained 95.3 % of the variance were retained. The number of principal components to use is normally not straightforward. We used seven to make a comparison with the seven variables obtained by the quantitation of important spectral ranges as described above. By chance, the seven principal components retained more than 95% of the variance, which is also a limit used to choose the number of principal components. After PCA, each voxel was described by 7 variables (called scores, figure 5.1d) obtained from the MRSI modality and four from MRI (figure 5.1e).

Exploration of the classes in feature space. In order to develop a classifier which is able to classify a voxel as belonging to one of the investigated classes or as 'unknown' or as 'undecided' between some tissue classes, we explored the distribution of each class in feature space and compared it with the distribution of all other classes. This was done for each class in both training sets created by the two data reduction methods. First, the data representing the voxels (objects) of the class under investigation was split, separating at random ten percent (A) from the rest (B). Then the covariance matrix of B was calculated and used to calculate the mahalanobis distances (d) of all objects in A with respect to the centroid of B. Also the mahalanobis distances of all objects in the other classes with respect to the centroid of B were calculated. This procedure was repeated successively leaving the next 10 percent of the class under investigation out, until all objects were left out once. This procedure was repeated for all classes. Eventually, for each tissue class a plot was created which showed the percentage of objects against the mahalanobis distance from the classes centroid for members of the class itself and for non members. An example of such a distribution plot is shown in figure 5.2. The solid line represents the percentage of healthy objects within a certain mahalanobis distance d from the centroid of the healthy class. 100% of the healthy objects in the training set

(created by the first data reduction method) are within a distance of 6 d from the healthy centroid. The dotted line represents the percentage of all other objects in the training set (not belonging to healthy) within a certain distance to the centroid of the healthy class. Only a small percentage of non healthy objects are within a distance of 6 d from the healthy centroid, and the healthy class is therefore relatively isolated from the other classes.

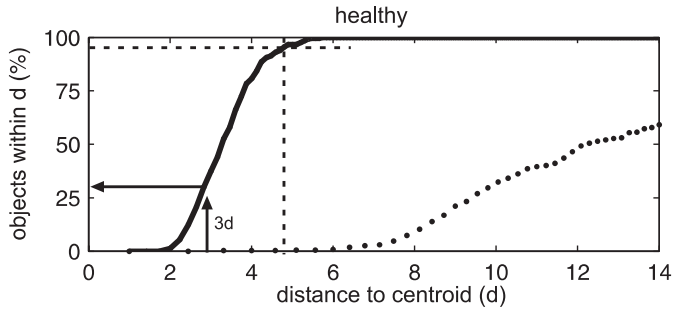


Figure 5.2. *Distribution plot showing the distribution of 'healthy' objects (solid) and objects from other classes (dotted) with respect to the centroid of the healthy class.*

Calculation of probabilities of class membership. The probability of class membership of a new object for a class was calculated based on the true distribution of the class, which is reflected by the distribution plot. The density at point $\mathbf{x_i}$ of a multivariate normal distribution (with p dimensions), characterized by the mean μ and covariance matrix $\mathbf{\Sigma}$ is given by:

$$\phi(\mathbf{x}_i|\mu,\mathbf{\Sigma}) = (2\pi)^{-p/2}|\mathbf{\Sigma}|^{-1/2}\mathrm{e}^{-(\mathbf{x}_i-\mu)^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{x}_i-\mu)/2} \qquad 1$$

In a non normalized way, the density is a direct estimator of the probability that an object at point $\mathbf{x_i}$ belongs to this multivariate normal distribution. If the mahalanobis distance of a new object to a class centroid is know, the density and thus probability of class membership can be calculated using this equation. In the proposed approach a slightly different method was used, based on the 'true' distribution of each class shown in the distribution plots.

To calculate the probability of class membership of a new object to the healthy class (for example), first its mahalanobis distance to the class centroid was calculated. Using the experimentally determined distribution plot in figure 5.2, the percentage of objects of the training set within this distance could be red. E.g., when an object has a distance of 3d from the centroid of the healthy class, then 28% of the objects in the training set was within this distance

(arrow at 3d in figure 5.2). Last, from a theoretical determined curve, which showed the number of objects a certain distance from the centroid of a multivariate normal distribution against the density, the density could be red. If the distribution plot in figure 5.2 was precisely following the theoretical distribution of a multivariate normal distribution, then the density calculated by our approach would give the same result as the density calculated by equation 1.

Classification of voxels. In order to use the distribution plots as a classifier, thresholds have to be set for each class. The value of the threshold is depending on the errors one is accepting to take for a specific tissue class. If one wants to classify 95% of the healthy objects correctly (the α error is 5%), then the threshold has to be set at the position shown in figure 5.2 (dashed vertical line). However, this means that ~2% of the objects will be classified incorrectly (β error) as healthy.

Classification of a new voxel is performed by projecting it in the feature space. Then, its distance to the centroid of each class is calculated. The voxel is only assigned to the closest class when its distance to the centroid does not exceed the threshold. If the distance exceeds the threshold, then it is classified as 'unknown'. If overlapping classes are found, rules can be setup to classify voxels as 'belonging to either one of these classes', as described in the discussion. If a voxel is classified as such, it may be re-classified with another classifier, like a support vector machine, especially constructed for the classification between two classes [25].

Evaluation of the system. Evaluation of the system was performed at four levels. First, the best data reduction method was selected by comparison of the distribution plots constructed with both methods. Secondly, the classifier was tested by a leave one out procedure on the training set (n= 524). Leave one out testing is an accepted validation procedure when the number of samples is low. The leave one out procedure was executed in the following way. From the training set the first sample was left out. The remainder of the training set was used to calculate new (independent to the left out sample) distribution plots. Then the left out sample was classified with use of the calculated distribution plots. The classification thresholds were kept the same during the leave one out procedure. This was repeated until all samples were left out of the training set once.

Third, the classifier was evaluated using all the data from all patients in the training set. This was performed by calculation of the classification maps and probability of classification maps. Most of the voxels of the patients were not used in the training set and therefore could be used for validation. Since no 'true' tissue type for these voxels were available, all maps were visually inspected for obvious misclassifications. Regions which could be assigned visually as healthy tissue, CSF or the tumor by means of the MR images had to be homogenously classified as such in the classification map. Last, the classification results of patient I-1283 (the meningioma) and patient I-1281 (for which no consensus was reached) were also used for validation.

Results and discussion

Exploration of classes in feature space. The distribution plots can be used to compare data reduction techniques as described above, or the influence of other processing steps, e.g. normalization or scaling. In this study only a comparison of data reduction techniques was made. Each subplot in figure 5.3 shows the distribution plot of a tissue class. The solid-black and dashed-gray visualize the percentage of objects of the class itself within a certain mahalanobis distance from its centroid. Theoretically, the squared mahalanobis distance should follow the Hotelling T^2 statistics [26], which is plotted as the gray thick line for each class. For all tissue classes the plots follow T^2 statistics, suggesting that the multivariate data is normally distributed in space and that using the mahalanobis distance is justified for both approaches. This is important, because the probabilities can only be calculated if the data follows a normal distribution. The approach using quantitated values seems to be better than the approach using PCA, since objects not belonging to a class are always further away from the class centroid for this method. This is especially visible for the grade III and grade IV class (figure 5.3d and 5.3e) in which the black-dotted line is lower than the gray dashed-dotted. This suggests that the classes are more separated in the approach using quantitated values and therefore this data set is used in the remainder of the paper.

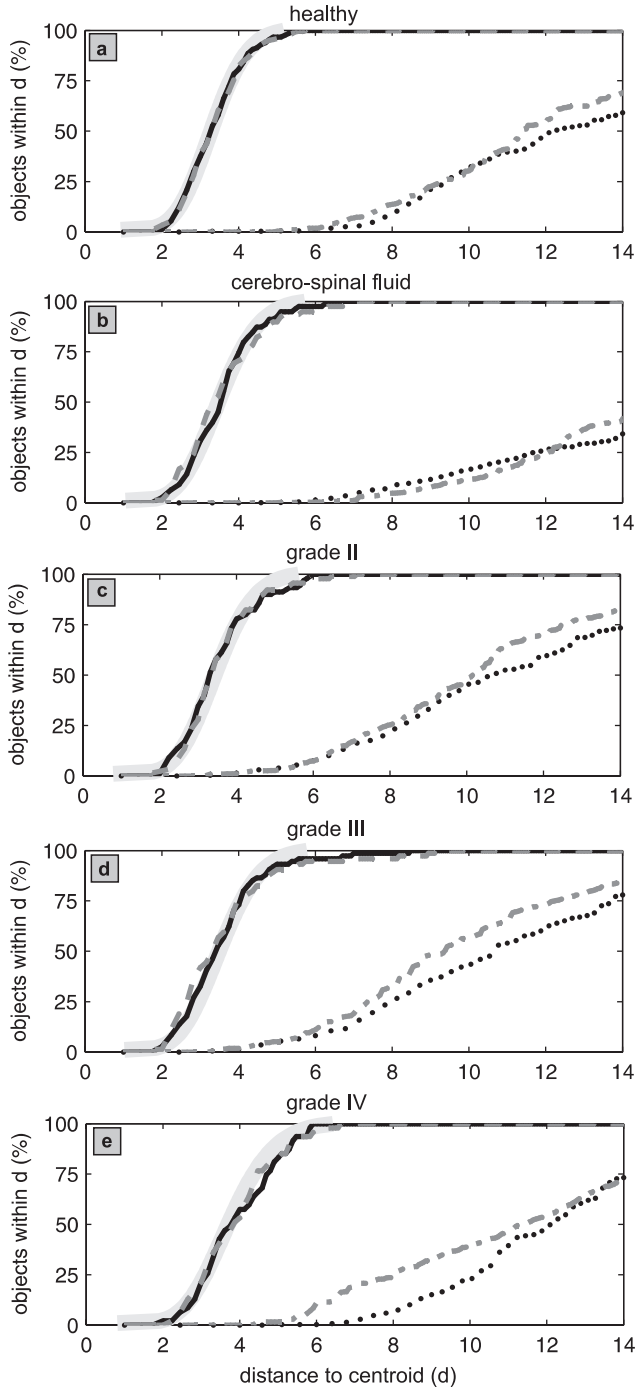


Figure 5.3. Distribution plots for each class using both data reduction approaches.

Caption of figure 5.3.
a) The distribution of the healthy class is plotted in solid-black for the quantitation of spectral ranges method, and in dashed-gray for the PCA method. Both distributions follow Hotelling T^2 statistics which is plotted as a thick gray band. Also the percentage of non-healthy training samples within a certain mahalanobis distance from the centroid of the healthy class is plotted for each data reduction approach. The dotted-black corresponds with the quantitation of spectral ranges method, and the dashed-dotted-gray for the PCA method. b, c, d, e) The same as a), but now for the CSF, grade II, grade III and grade IV class, respectively.

To investigate the distribution of all tissue classes which surround a particular class, we plotted the mahalanobis distance of all objects (including MNG) within a class to the centroid of the particular class under investigation (figure 5.4). The surrounding of the healthy class is relatively empty. At a mahalanobis distance of 6 d from its centroid, objects from the cerebro-spinal fluid class arise. Objects from other classes however, are almost absent until a distance of 8 d. The CSF class has some overlap with objects from the healthy class at 6 d, but no overlap with other classes at all (figure 5.4b). This suggests that healthy tissue and CSF are dense clouds in feature space, which have some overlap with each other, but not with malignant tissue. Therefore, one may expect some error between these classes, but low error between healthy/CSF and any of the other classes. This is in agreement with findings of other researchers [6, 12, 13] who found that healthy tissue can be classified with a high reliability. This can be explained by the stringent concentrations of metabolites in the human brain for healthy tissue. The overlap of CSF with healthy is likely due to partial volume effects.

The grade II class overlaps with the grade III class, but not with grade IV or MNG (figure 5.4c). It is therefore very well possible to separate grade II from grade IV and MNG objects, but separation of grade II from grade III will be hard. The same can be concluded from figure 5.4d, which describes the grade III distribution. The grade IV class is separated from all other classes, suggesting that this class can be classified with a high reliability. As pointed out, we included the MNG class in figure 5.4a-d, because it gives relevant information about the location of MNG in feature space.

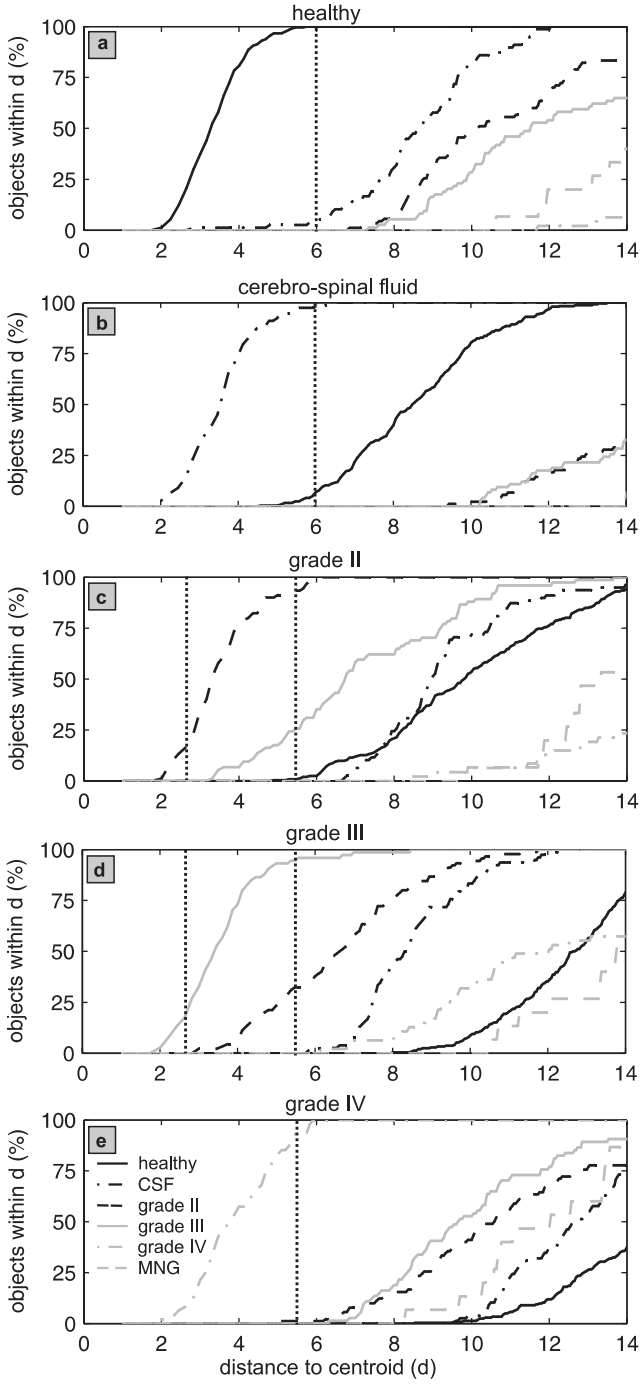


Figure 5.4. Distribution plots of each class and its surrounding classes using the quantitation of spectral ranges method for data reduction.

Caption of Figure 5.4
In each sub-plot the solid black line represent the percentage of healthy objects within a certain distance to the centroid of the class under investigation. The dashed-dotted-black line represents CSF, the dashed-black line, grade II. The solid-gray, dashed-dotted-gray and dashed-gray respectively represent grade III, grade IV and MNG. The vertical dotted lines indicate the thresholds used for the classification.

The distribution plots are very import for the interpretation of the classes in feature space. If a new class is added, one can directly see whether this new class has a significant overlap with any of the other classes. Assume that a new class A is overlapping with class B. If a new object is classified to class A, but the probabilities for both class A and B are high, then the reliability of classification is low. The object might also be a member of class B. If the classes A and B are from totally different tumor types, which normally do not occur together in the brain, then the classification is A, with a low reliability. On basis of other information the clinician might decide that the 'true' classification is A. It can also be that the classes A and B are from the same tumor type, but with a different grade. This is for example the case for the grade II and grade III classes of glial tumors described in this paper. A mixture of these two tumor types can occur in the brain. If the probability for class A and B are high now, and the two classes are from different grades, then the tissue could also be a mixture of both classes and thus heterogeneous. The expert user has to keep this in mind during the usage of the classifier.

As explained in the methods section, probability maps for a new patient can be constructed using the distribution plots of figure 5.4. Maps are presented from patients with a histologically proven grade III (figure 5.5a and d), a grade II (figure 5.5b) and a grade IV (figure 5.5c) tumor. The darker a voxel in a map, the higher the probability that the underlying tissue belongs to the class represented by the map. The images give a very consistent indication of the different tissues in the brain, indicating robustness of the system. Remember that only several voxels of each patient have been used in the training set. The number of voxels used in the training set and the number of total voxels is specified in the figure caption for each patient. For the interpretation of the probability maps, it is important to compare them region wise, because the probabilities are relative. When a region in the probability map of grade II is dark, while it is bright on the other maps, then the estimation of class membership is high. When the region is dark in more than

one map, then the estimation of class membership is lower. As explained, the probabilities of classification might also be important for the assessment of tumor heterogeneity. When the probability in a region is high for only one tissue type, then the tissue will be, in general, homogeneous. If the probability in a region is high for two tissue types, then the region might be a mixture of these tissues, and therefore heterogeneous. However, this can only be the case if the mixture of these two tissues are possible from a clinical point of view. Otherwise, on can only say that the reliability of classification is low. In addition to other classification procedures which try to assess tissue heterogeneity with a spatially resolved classifications algorithm [4, 6] or the construction of metabolic maps [27, 28], the use of statistically calculated probabilities could be an important asset to the mapping of tumor heterogeneity.

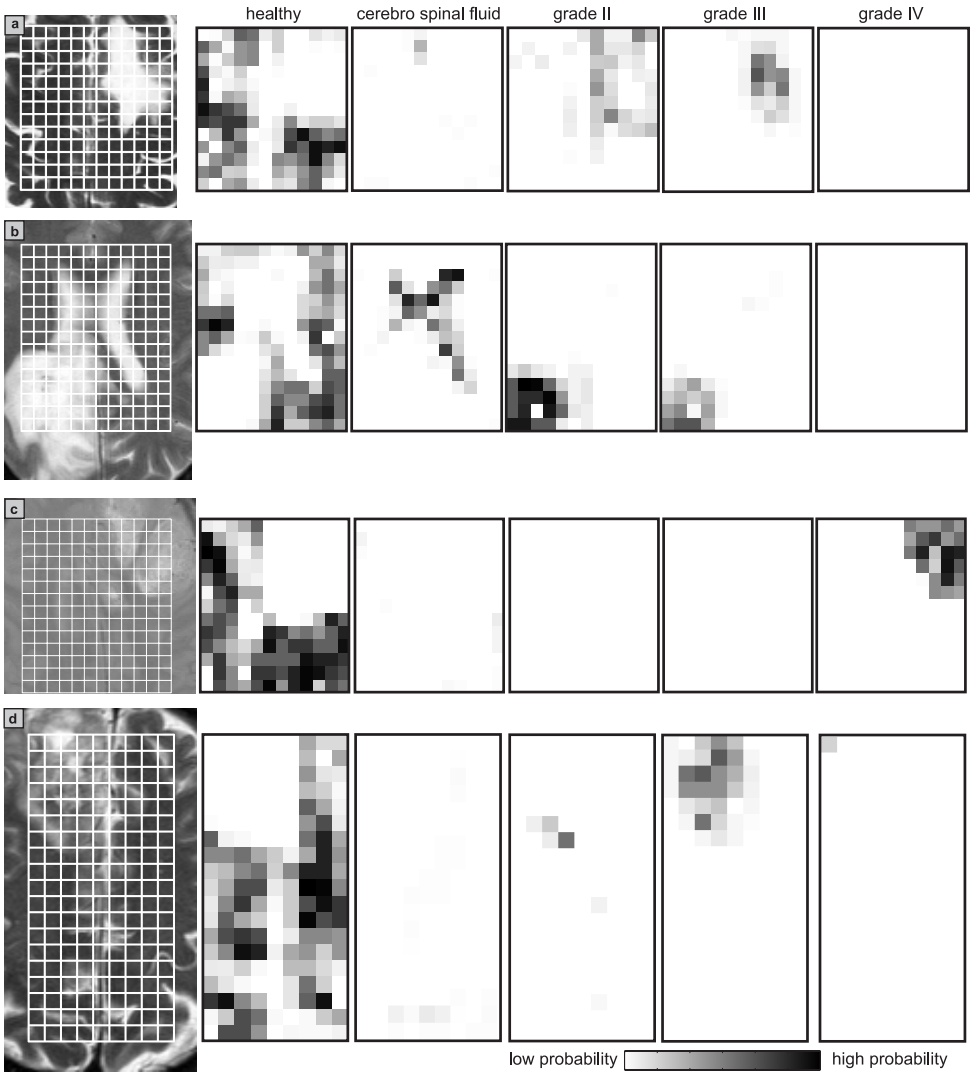


Figure 5.5. Overview of probability maps of 4 patients. *For each patient a MR image is shown on the left with a contrast that clearly identifies the tumorous region. The MR images are zoomed and only the MRSI grid within the STEAM box is depicted. Next to the MR image the probability maps for the five classes are shown. a) patient I-1229 (6/156), which has a grade III tumor. After the patient label, the number of voxels used in the training set and the total number of voxels for this patient are depicted between brackets. The MR image is T_2-weighted. b) patient I-1275 (21/180), which has a grade II tumor. The MR image is T_2-weighted. c) patient I-1285 (55/168), which has a grade IV tumor. The MR image is proton density-weighted. d) patient I-1212 (49/171), which has a grade III tumor. The MR image is T_2-weighted.*

Classification of voxels. For the classification, we created classification rules based upon the results presented in figure 5.4. The healthy and the CSF classes are isolated in feature space, having some overlap with each other. To minimize the α error, we have set the threshold to 6 d for these classes as shown by the vertical dotted line in figure 5.4a and 5.4b. The same can be concluded for the grade IV class, with a threshold of 5.5 d. This leaves some α error for grade IV, but minimizes the chance that a healthy or CSF object will be falsely classified as grade IV. Because of the overlap of the grade II and grade III classes we defined different rules for these classes. When an object is very close ($\leq$ 2.5 d) to the centroid of one of these classes it is appointed to it (threshold at 2.5 d), but when the distance is between 2.5 d to 5.5 d, other rules were applied. a) If the object was closest to the grade II class and the distance of the object to the grade III class was 1.3 times larger than to the grade II class, then it was assigned to the grade II class. This rule can be applied because all other classes are far away from the grade II class. b) If the distance of the object to the grade III class is smaller than 1.3 times the distance to the grade II class, then it is a grade II or a grade III member and thus 'undecided' between *these* two classes. We applied identical rules when the object was closest to the grade III class. The value of 1.3 was found by repeating the classification for different values, and assessing the number of objects which were classified as 'undecided'.

Classification using leave one out is excellent for the healthy (100 % correct), CSF (97 % correct, the 3 % being misclassified as healthy) and grade IV (100 % correct) class. Classification results for GII and GIII were promising. 83% of the GII objects were classified correctly and 88% of the GIII class. However, the largest part of the misclassified objects in the GII and GIII classes were classified as 'undecided' (12 % for the GII and 4 % for the GIII class), which is to be expected since these classes are overlapping. 2% of the GII objects were misclassified as GIII and 3% of the GIII as GII, respectively. Last, some objects of the GII and GIII classes were classified as 'unknown' (3% and 5%). All MNG objects were classified as 'unknown', which is correct as we did not build the classifier to be able to take MNG into account.

To compare the method with a commonly used algorithm [4, 6], we performed linear discriminant analysis (LDA) [24] with use of leave one out. In general, the number of correctly classified objects for each class using LDA was in correspondence with the presented method, but when LDA was used, the misclassified objects were often assigned to classes not expected. Objects from the healthy and CSF classes were wrongly assigned to grade II (respectively 0.5 % and 1 %). 5 % of the GII were misclassified as GIII, and 11 % of the GIII as GII. 1.5 % of the GIII were misclassified as GIV.

The results of LDA may be caused by the fact that LDA can only be used if the classes under investigation have equal covariance in size and shape. In our data set this was not the case. The healthy and CSF classes are much more condensed than the tumor classes. This will probably always be the case, because the highly structured brain will give rise to spectra with low variance, while heterogeneous tumor tissue with a broader range of morphologic and metabolic abnormalities will give rise to MR images and spectra with high variance.

In figure 5.6, the classification results of the patients presented in figure 5.5 are shown, together with two other patients. The patient in figure 5.6e (I-1318) has a grade II tumor, the pathology of the tumor of the patient in figure 5.6f (I-1281) is unclear, since the pathologists who validated the tumor disagreed between grade II and grade III. Therefore this patient has not been used during training of the classifier, but was used as a test case. Also all other patients have been evaluated, showing similar results as the patients presented.

Although only a small part of the voxels from patients a to e are used for training the classifier, the results look very good. As an example we have indicated the voxels used in the training set for the patient in figure 5.6a with white circles. Only 6 of the 156 voxels of this patient were used for training! Important considerations for the evaluation are; a) voxels which are identically classified tend to be spatially grouped, b) the spatial distribution of classification is in agreement with morphological information from the MR images, c) in most patients only one tumor type is assigned, and this assignment is always in agreement with the pathology of the patient, d) most grade III patients also show grade II voxels which can be expected, because the pathologist types the patient with respect to the most malignant tissue present. However, grade II patients do not show grade III classified voxels. This is important since otherwise a clinician could assume grade III tumor based on 1 or 2 falsely classified grade II voxels. In figure 5.6d and e, voxels

have been misclassified as grade IV. In the case of d the misclassification could have occurred because it is a corner voxel. These voxels normally have a low signal, since the boundary setup by the STEAM box is never perfect.

The classification result of the patient in figure 5.6f is fairly good for the healthy and CSF tissue. There are misclassified voxels in the border region of healthy and CSF which might be caused by partial volume. In the tumor area none of the voxels have been classified as grade III, suggesting this is a grade II tumor, but the probability of grade II shown on the probability maps (inset) is not very high for most voxels. The disagreement between the two radiologists of this patients pathology is therefore reflected in the results of the classifier and probability maps.

Last, the patient with the meningioma tumor was evaluated (classification result not shown). Again, no voxels from this patient were selected for the training set. The classification map shows normal tissue in regions outside the tumor. The area of the tumor is classified as 'unknown', which is correct as the classifier was not build to be able to classify meningiomas.

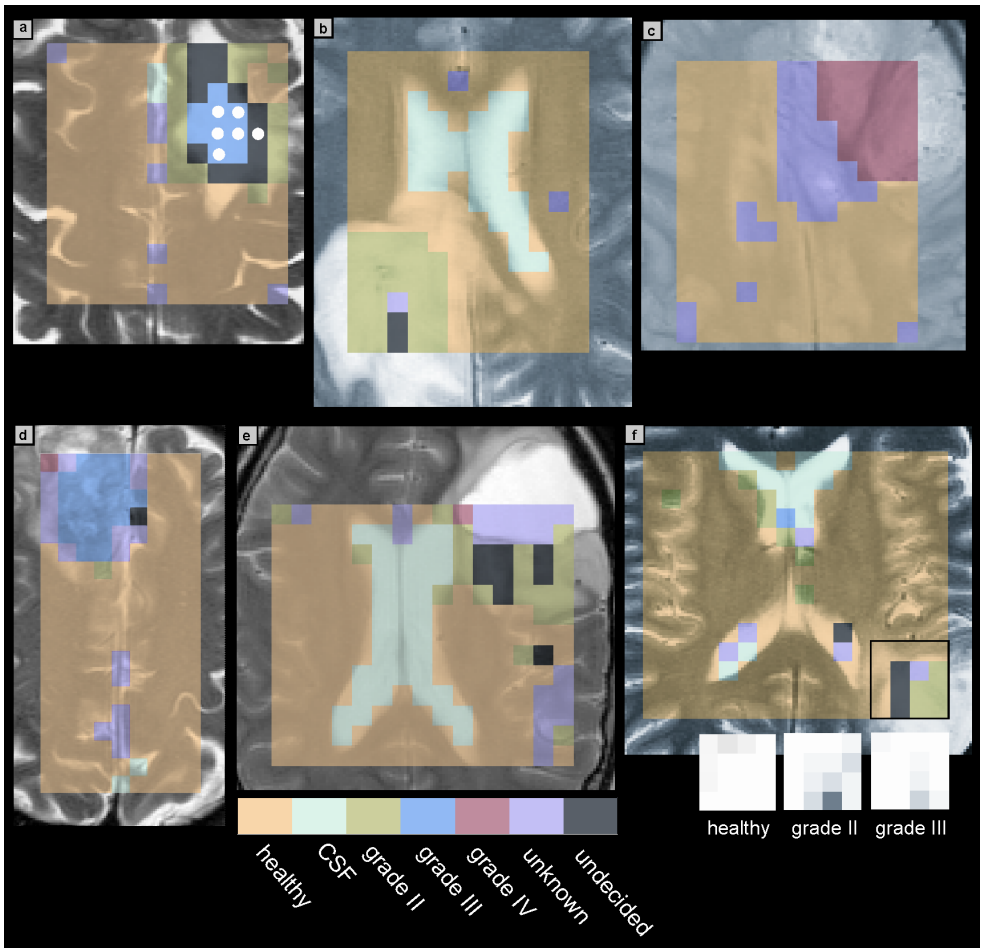


Figure 5.6. Classification result of six patients. *The classification is projected over a MR image with a contrast that clearly identifies the tumorous region. a) patient I-1229, pathology: grade III, T_2-weighted image, b) patient I-1275, pathology: grade II, T_2-weighted image, c) patient I-1285, pathology: grade IV, proton density-weighted image, d) patient I-1212, pathology: grade III, T_2-weighted image, e) patient I-1318 (42/195), pathology: grade II, T_2-weighted image, f) patient I-1281, no consensus of the patients pathology was reached. The radiologists differed between grade II and III. None of the voxels of this patient have been used for training. The inset shows a section (indicated by the black square on the classification map) of 3 probability maps of healthy, grade II and grade III. Note that a voxel classified as 'undecided' can either be a grade II or a grade III.*

Conclusions

A new classification approach was developed which enables the classification and probability of classification of brain tumors on the basis of magnetic resonance spectroscopy and imaging data. The model can be used to determine the best preprocessing or data reduction technique in feature space and to recognize overlap of classes. The results of classification, including probabilities, are presented as images to facilitate the interpretation in a clinical environment.

Comparison of PCA and quantification of spectral ranges to reduce the number of variables in feature space gave similar results. The last method was favored, because it was slightly better and because the variables retain a meaningful property. However, PCA is much easier to perform and is not subjective to quantification error and is therefore a good competitor.

Exploration of the classes in feature space as presented is a powerful method to compare class distances, distributions and class boundaries. It can be used to identify tumor classes which are not separable, because they overlap in feature space. The overlap between the grade II and III tumor classes can be understood if the method to identify the pathology is considered. When only a small part of the tumorous cells are found to be grade III, the pathology of the patient will be grade III, whereas the spectral features may arise from a bulk of grade II tissue. It is therefore never possible to separate these two classes perfectly. The classification thresholds of the model thus have to be set to minimize the number of voxels falsely classified as grade III in patients with a grade II tumor.

The calculation of probabilities of classification from MRI and MRS data is relatively new. The probability maps are very important for the interpretation of the classification image. It provides information of the reliability of classification for each voxel, shows whether voxels have a high probability for a class they were not classified to and can be used as a direct identifier of tissue heterogeneity in certain cases.

Unfortunately, as in any classifier, the classification approach is limited to classification of the classes in the training set. There are many different types of tumor, so an addition of classes to the training set in future would be highly desirable. Then it would also appear whether totally different tumor classes show overlap in feature space. However, the presented approach recognizes 'unknown' tumor classes, so at the moment it is a valuable method to differentiate between glial tumors and other tumors. Furthermore, it provides the grade and probability of class membership for glial tumors, which are important parameters in tumor diagnosis.

Acknowledgment

The authors acknowledge Pieter Wesseling (PhD, MD) and Janneke Schuuring (MD) for histopathological identification and selection of patients. The EU funded INTERPRET project is gratefully acknowledged (IST-1999-10310; http://carbon.uab.es/INTERPRET/).

References

1 Noworolski, S. M.; Nelson, S. J.; Henry, R. G.; Day, M. R.; Wald, L. L.; Star-Lack, J.; Vigneron, D. B. *Magnetic Resonance in Medicine* **1999**, *41*, 21-29.

2 Govindaraju, V.; Young, K.; Maudsley, A. A. *NMR in Biomedicine* **2000**, *13*, 129-153.

3 Rijpkema, M.; Schuuring, J.; van der Meulen, Y.; van der Graaf, M.; Bernsen, H.; Boerman, R.; van der Kogel, A.; Heerschap, A. *NMR in Biomedicine* **2003**, *16*, 12-18.

4 Preul, M. C.; Caramanos, Z.; Collins, D. L.; Villemure, J.; Leblanc, R.; Olivier, A.; Pokrupa, R.; Arnold, D. L. *Nature Medicine* **1996**, *2*, 323-325.

5 Mader, I.; Roser, W.; Hagberg, G.; Schneider, M.; Sauter, R.; Seelig, J.; Radue, E. W.; Steinbrich, W. Magnetic resonance materials in physics, biology, and medicine (*MAGMA)* **1996**, *4*, 139-150.

6 Szabo De Edelenyi, F.; Rubin, C.; Estève, F.; Grand, S.; Décorps, M.; Lefournier, V.; Le Bas, J.; Rémy, C. *Nature Medicine* **2000**, *6*, 1287-1289.

7 Nelson, S. J.; Huhn, S.; Vigneron, D. B.; Day, M. R.; Wald, L. L.; Prados, M.; Chang, S.; Gutin, P. H.; Sneed, P. K.; Verhey, L.; Hawkins, R. A.; Dillon, W. P. *Journal of Magnetic Resonance Imaging* **1997**, *7*, 1146-1152.

[8] Hagberg, G. *NMR in Biomedicine* **1998**, *11*, 148-156.

[9] El-Deredy, W. *NMR in Biomedicine* **1997**, *10*, 99-124.

[10] Somorjai, R. L.; Nikulin, A. E.; Pizzi, N.; Jackson, D.; Scarth, G.; Dolenko, B.; Gordon, H.; Russell, P.; Lean, C. L.; Delbridge, L.; Mountford, C. E.; Smith, I. C. P. *Magnetic Resonance in Medicine* **1995**, *33*, 257-263.

[11] Meyerand, M. E.; Pipas, J. M.; Mamourian, A.; Tosteson, T. D.; Dunn, J. F. *American Journal Neuroradiology* **1999**, *20*, 117-123.

[12] Minguillón, J.; Tate, A. R.; Arús, C.; Griffiths, J. R. *Lecture Notes in Computer Science* **2002**, *2364*, 282-292.

[13] Tate, A. R.; Griffiths, J. R.; Martinez-Pérez, I.; Moreno, A.; Barba, I.; Cabañas, M. E.; Watson, D.; Alonso, J.; Bartumeus, F.; Isamat, F.; Ferrer, I.; Vila, F.; Ferrer, E.; Capdevila, A.; Arús, C. *NMR in Biomedicine* **1998**, *11*, 177-191.

[14] Hagberg, G.; Burlina, A. P.; Mader, I.; Roser, W.; Radue, E. W.; Seelig, J. *Magnetic Resonance in Medicine* **1995**, *34*, 242-252.

[15] Howells, S. L., Maxwell, R. J.; Peet, A. C.; Griffiths, J. R. *Magnetic Resonance in Medicine* **1992**, *28*, 214-236.

[16] Soher, B. J.; Young, K.; Govindaraju, V.; Maudsley, A. A. *Magnetic Resonance in Medicine* **1998**, *40*, 822-831.

[17] Bartha, R.; Drost, D. J.; Williamson, P. C. *NMR in Biomedicine* **1999**, *12*, 205-216.

[18] Clarke, L. P.; Velthuizen, R. P.; Camacho, M. A.; Heine, J. J.; Vaidyanathan, M.; Hall, L. O.; Thatcher, R. W.; Silbiger, M. L. *Magnetic Resonance Imaging* **1995**, *13*, 343-368.

[19] International network for pattern recognition of tumors using magnetic resonance

[20] http://carbon.uab.es/INTERPRET/interpret_home.shtml

[21] Frahm, J.; Merboldt, K. D.; Hanicke, W. *Journal of Magnetic Resonance* **1987**, 72, 502-8

[22] Simonetti, A. W.; Melssen, W. J.; van der Graaf, M.; Heerschap, A.; Buydens, L. M. C. *Journal of Magnetic Resonance* **2002**, *159*, 151-157.

[23] Pijnappel, W. W. F.; van den Boogaart, A.; de Beer, R.; van Ormondt, D. *Journal of Magnetic Resonance* **1992**, *97*, 122-134.

[24] Vanhamme, L.; Fierro, R. D.; Van Huffel, S.; de Beer, R. *Journal of Magnetic Resonance* **1998**, *132*, 197-203.

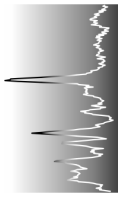
[25] Vandeginste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics: Part B*; Elsevier, 1998; Chapter 33.

[26] Lukas, L.; Devos, A.; Suykens, J. A. K.; Vanhamme, L.; Van Huffel, S.; Tate, A. R.; Majós, C.; Arús, C. *ESANN'2002 proceedings - European Symposium on Artificial Neural Networks* 2002, 131-136.

[27] Candolfi, A.; De Maesschalck, R.; Jouan-Rimbaud, D.; Hailey, P. A.; Massart, D. L. *Journal of Pharmaceutical Biomedical Analysis* 1999, *21*, 115-132.

[28] Segebarth, C.; Balériaux, D.; Luyten, P. R.; den Hollander, J. A. *Magnetic Resonance in Medicine* 1990, 13, 62-76

[29] Furuya, S.; Naruse, S.; Ide, M.; Morishita, H.; Kizu, O., Ueda, S.; Maeda, T. *NMR in Biomedicine*, 1997, 10, 25-30



Chapter 6

A medical decision support system for brain tumor diagnosis exploiting the combined power of proton magnetic resonance imaging and spectroscopic imaging data

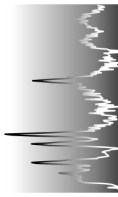
AW Simonetti[♣], *M van der Graaf*[♦], *WJ Melssen*[♣], *P Wesseling*[♠], *GJ Postma*[♣], *J Schuuring*[♥], *ELM Mooyaart*[♦], *LMC Buydens*[♣], *A Heerschap*[♦]

[♣] **Laboratory for Analytical Chemistry, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands**

[♦] **Department of Radiology, University Medical Center Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands**

[♠] **Department of Pathology, University Medical Center Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands**

[♥] **Department of Neurology, Groene Hart Hospital, P.O.Box 1098, 2800 BB Gouda, the Netherlands**



Abstract

Purpose. To investigate whether an automated medical decision support system that uses pattern recognition and the combined power of magnetic resonance imaging and proton spectroscopy can aid radiologists in the diagnosis of brain tumors.

Materials and methods. Information is extracted from the MR data and used to segment pathologically similar brain regions. A classification model classifies the segments and individual voxels. For each classification result a probability is calculated.

Data from 22 subjects is used to build and assess the system. Evaluation is performed by the comparison of classification results from the segmentation and individual voxels, by visual evaluation and by a clinical evaluation by four expert radiologists.

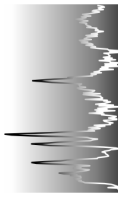
Results. The clinical evaluation shows that radiologists agree with the location and extent of brain tumor segments, as found by the segmentation. Of 16 brain tumors, 15 were correctly classified. The probability of classification for individual voxels are mostly high in the center of the region that is considered to be the tumor.

Conclusion. The support system is a user friendly tool, that provides information on the location, type and grade of brain tumors. The reliability of classification can be assessed, and all results are presented as images to facilitate an easy interpretation. The large amount of data obtained by the combination of MR imaging and spectroscopic data can be exploited by automated pattern recognition to achieve a robust tumor classification.

Introduction

In the diagnosis, therapy prediction, and assessment of brain tumors, imaging techniques play a critical role, among which magnetic resonance imaging (MRI) is most prominent. Since the assessment of tumor tissue in the brain, based on imaging procedures alone is still a challenging problem, proton Magnetic Resonance Spectroscopy (MRS) has been introduced as an additional noninvasive diagnostic tool. Proton MRS provides completely different information as compared to MR imaging, but can be obtained with most 1.5 T MRI instruments used for routine radiological examinations. MRS spectra provide localized metabolic information, related to cell membrane proliferation, neuronal damage and energy metabolism. MRS can be extended to an imaging mode called MRS imaging (MRSI), in which the acquisition of multiple voxels provides spatially encoded chemical information. MRS(I) has been found useful in the diagnosis and grading of brain tumors[1, 2] and could possibly reduce the need for an invasive biopsy[3-6].

The application of different MR approaches results in scattered information within a large amount of data, from which it may be difficult to derive a final diagnosis. To avoid that the clinician is overwhelmed with information, automated statistical methods have to be developed to process, combine and classify the data. In the past, several investigators have exploited MRI or MRSI[7] information to classify brain tumors[1, 3, 5, 6, 8-11]. Only a few researchers[4, 12] combined the spectral and image features to improve the classification result. If MRSI is performed next to MRI, then the combination of data from both modalities seems to be a logical extension, based on the idea that this synergy of data will improve substantially the power and robustness of the classification model.



Next to classification, also the exact boundary and degree of heterogeneity of the tumor is very important in the diagnosis, prognostics and treatment of brain tumors. The use of metabolic maps, derived from MRSI data has been discussed frequently in literature[13-17]. These maps provide a good insight in the increase or decrease of important brain metabolites, like creatine, choline and N-acetylaspartate. Often, a significant statistical change in one of the metabolite levels is used as a marker for malignancy of the tumor. However, in most cases merely univariate methods are used, while the availability of information is highly suited for a multivariate approach.

In order to increase the reliability of classification, several classification approaches have been developed. The tree-like classifiers[18] use multiple binary classifiers, each optimized for a specific decision. Another approach is to use a so-called consensus diagnosis[9], in which several independent classifiers work in parallel. Last, reliabilities can be estimated from the calculation of posterior probabilities[11] or from probabilities based on the actual distribution of the data itself[12].

This paper reports on the development of a decision support system to investigate whether multivariate automated pattern recognition of combined MRI and MRSI data could aid the clinician in diagnosing brain tumors. The methodology provides a spatially resolved classification result, a reliability of classification and a consensus diagnosis, that could lead to a better understanding of brain tumor type and morphology.

Materials and methods

This study has been approved by the ethical committee of the university medical center and from all patients written informed consent was obtained. MR data was acquired following a well defined acquisition protocol designed within the EU project INTERPRET[19]. Examinations were performed on a 1.5 T Siemens Vision whole body system, using a CP-head coil. First, conventional transversal T1-weighted, T2-weighted and proton density weighted images (PD-image) were acquired. The MR images were collected using the following parameters: 23 slices of 5-mm thickness, inter-slice distance 1.5 mm, FOV 230 mm, matrix 256x256. For the T1-weighted images (TE/TR=15/644 ms) a standard spin echo sequence was used. The PD-images (TE/TR= 16/3100 ms) and T2-weighted images (TE/TR= 98/3100ms) were simultaneously acquired using a turbo spin echo sequence. This was followed by acquisition of a set of T1-weighted images after intravenous administration of a bolus injection of 15 ml 0.5 M Gd-DTPA Gadolinium-DTPA (Gd-image).

The protocol was concluded by water suppressed and unsuppressed proton MR Spectroscopic Imaging (1H-MRSI) with the use of a 2D STEAM 1H-MRSI sequence with short echo time. The STEAM box was positioned in a transversal plane through the brain showing the largest Gd enhanced tumor area on a Gd-image. It was placed entirely in the brain parenchyma avoiding leakage of disturbing signals from fatty tissue surrounding the skull. MRSI parameters were: 16x16x1024 samples, TR/TE/TM = 2000 or 2500 ms / 20 ms / 30 ms, slice thickness = 12.5 or 15 mm, FOV = 200 mm, spectral width = 1000 Hz and NS = 2 (1 for the MRSI without water suppression).

Data sets. Data of a total of 18 patients and 4 volunteers were used in the study. Each patient case had passed strict quality control and validation procedures, including consensus histopathologic determination. 15 patients were diagnosed with a glial brain tumor of specific grade and one with a meningioma. For the two other patients no consensus of the grade could be reached, but both had a glial tumor. For the set-up of the training set, the data of the 15 patients diagnosed with a glial tumor of specific grade were divided into three classes according to the rules of the World Health Organization (WHO): grade II (6 cases), grade III (4 cases) and grade IV (5 cases) tumors. In order to create a sufficiently large training set, multiple MRSI voxels located within the tumor were selected from each patient. Also classes for healthy brain tissue and cerebrospinal fluid (CSF) were selected. The data for healthy tissue were selected from the contra-lateral brain region of several patients and from 4 healthy volunteers. Data representing CSF were selected from CSF voxels which were not in close contact with the tumor region.

The patient with the meningioma was not used in the training set, but as a test case for contra-validation, that is, for this case the classifier should not assign the meningioma region to any of the classes used to construct the classification model.

Preprocessing of spectroscopic data. The MRI and MRSI data were preprocessed in the same unique way for each case and automated where possible. Preprocessing of the MRSI data[12] was performed as described in chapter 5 and comprised of the steps presented in figure 6.1. After preprocessing, each MRSI time domain signal was Fourier transformed to obtain a spectrum from which the real part of the region between 0.5 to 4.0 ppm was retained (230 spectral points). To normalize the signals originating from different patients, all spectra of one patient were divided by the mean water signal of all voxels within the patients STEAM box.

Finally, from the processed MR spectra, intensities of seven spectral ranges encompassing signals of important brain metabolites were determined by integrating the intensities within a window of 0.13 ppm. The integration regions were chosen around the expected chemical shift of the main resonances of the following metabolites; glutamate/glutamine (at 3.75 ppm), myo-inositol (3.56 ppm), choline (3.20 ppm), creatine (3.02 ppm), N-acetyl aspartate (NAA, 2.02 ppm), lactate/fatty acids (1.33 ppm) and fatty acids (0.90 ppm).

Matching image resolution. The MR images were aligned with respect to each other, by successively shifting the T1- weighted and Gd-image with respect to the PD-image until a maximum of spatial correlation was reached. Then, the image pixels not within the boundary of the STEAM box were discarded, and the resolution of the remaining part of the MR images was lowered to the resolution of the MRSI grid. This was performed by averaging the intensities of the image pixels which were covered by each spectroscopic voxel. Last, the low resolution images were scaled. After scaling, the minimum and maximum intensity values in an image were in agreement with the minimum and maximum values of the integrated regions of the spectral data. However, the minimum intensity value could not be lower than 0. After preprocessing, each voxel within the STEAM box was represented by 7 spectral variables obtained by the integration of metabolite regions and 4 variables extracted from the four MR images.

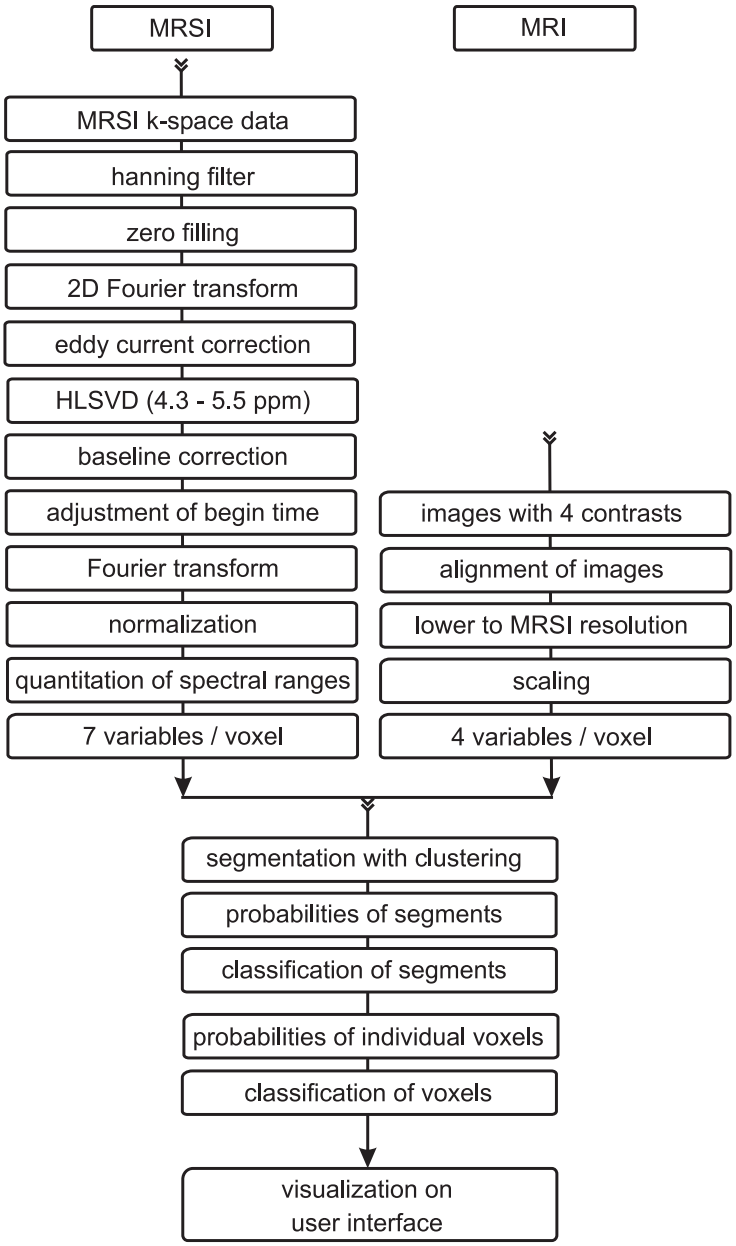


Figure 6.1. Flow diagram of the processing of data, followed by segmentation, classification and visualization procedures. *For specific settings[12] used for the processing steps, and general information about the eddy current correction[20, 21] used, HLSVD filtering[22] and baseline correction[23], we refer to chapter 3, 5 and 7.*

Segmentation with mixture modeling. For the segmentation of the brain, a clustering program based on mixture modeling was used. For an extended explanation of this method[7], we refer chapter4 and[24]. Mixture modeling fits the input data space with a set of multi-dimensional Gaussian bells, whose location, shape, orientation and volume have to be determined. The method starts with an initial partitioning of the data in as much fractions as the number of clusters (k) one expects. On basis of this partitioning, the parameters for the k multi-dimensional Gaussians are calculated. Then the conditional probabilities of all objects for all classes k are calculated. On basis of the probabilities of each object for all classes, the data can be re-partitioned and new parameters for the Gaussians can be calculated. This is repeated iteratively until convergence of the likelihood occurs. The likelihood is used as an estimator for the underlying distribution describing the spread of the data in an optimal way. The clustering can be repeated with differences in the model to find the optimal clustering. The most important difference concerns the number of clusters (Gaussians) one assumes to be present in the data.

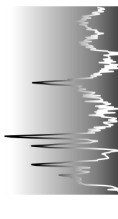
To select which clustering is optimal, the quality of the clustering can be assessed. The likelihood of the clustering is a first indication, but fails to incorporate the complexity and number of clusters of the model. A measure which corrects for this is the Bayesian Information Criterion (BIC)[25]. With the use of the BIC value, the model that fits the data best, if a specific number of clusters is used, is selected. Selection of the best number of clusters is left to the clinical expert to allow a synergy between the statistically based segmentation and the medical expertise (partially based on a patient's casus).

Practically, all voxels of a patient, using the 11 selected variables as input, are clustered. Clustering is repeated with 3, 4, 5, and 6 clusters, each with 6 different models (restrictions on the Gaussians). The clusterings with the best model for the 3, 4, 5, and 6 cluster solution are chosen automatically using the BIC values. Finally, the clinician decides what number of clusters is optimal for the best possible segmentation. At this point it is also possible to manually rearrange the segmentation by deselecting voxels or changing segment membership. The clustering result is projected on a high resolution image and the clusters can be colored according to the classification with the classification algorithm.

Classification. Previously, a classifier has been developed to classify grade II, grade III, and grade IV glial tumors, healthy tissue and CSF[12]. It classifies new objects to one of these classes, as 'unknown' (meaning that the object does not belong to any of the pre-defined classes) or as 'undecided' (indicating that the object possesses a multiple class membership) between grade II and grade III tumor. Classification is based upon the Mahalanobis distances[26] of a new sample towards the centroids of all classes. The sample is assigned to the class it has the closest distance to, if this distance does not exceed a critical value. If the distance exceeds this critical value, the sample is assigned as 'unknown'. Because the grade II and grade III classes were found to be overlapping, a sample is only assigned to one of these classes if its distance is sufficiently lower than to the other class. If the distance to both classes is within the same range, then the sample is assigned to be 'undecided' between grade II and III.

Next to classification, also relative probabilities of classification are calculated by the method. This is based on the distance of the sample to each class and the distribution of the classes. The probabilities of classification indicate whether the classification is reliable. If the probability for only one class is high, then the reliability of classification is high. If probabilities for several classes are in the same range, then the reliability of classification is low.

The classifier has been extended to facilitate the classification of segments found by the clustering model. This is performed by averaging the MRI and MRSI information of voxels belonging to the same segment, followed by classification of this average information. Theoretically, this is possible, if the voxels within a segment are from the same type of tissue. Classification could even become more robust, since the averaging of similar voxels might improve coherence of the input data.



Evaluation. To evaluate the decision support system and the classification results, the classification of the segments was compared with the classification of the individual voxels and with the results of the histopathologic determination. The classification and probability of classification maps were visually inspected to investigate whether the classification result showed a homogeneous section in the brain. Also the occurrence of outliers was investigated visually. Four radiologists of the University Medical Center Nijmegen (UMCN) evaluated the segmentation by assessing the results of five patients. For each case they had to draw the malignant area (or areas) in the patients brain, based upon all available image information and clinical information. After drawing the malignant area, the radiologists were asked to select from the 3 to 6 segmentation result, produced by the clustering the one which corresponded best to their drawing. Then they were asked to what extent the selected segmentation agreed with their drawing.

Results

The result of the classification strategy, that can be displayed on a monitor, is shown in figure 6.2. Figure 6.2a provides the clinician with the MR images of the patient. In figure 6.2b the segmentation of the brain into 3, 4, 5 and 6 segments respectively, is shown. Note that voxels assigned to different clusters in one segmentation solution may be combined in another clustering solution. The colors in each segmentation are only used to visualize the results, so corresponding colors for different segmentations have no specific meaning. After segmentation, the information from voxels that are within the same segment is averaged, and can be used for classification.

To help the clinician to decide which segmentation is best, figure 6.2c is displayed. It shows for each segment in the 3-6 segment solution of figure 6.2b, the probabilities that it belongs to one of the investigated tumor classes. Each row shows the probabilities of class membership of a specific segment, for all classes. Each cell is colored gray, a darker color means a higher relative probability for a class. The first row, for example, represents the blue segment in the 3 cluster solution. It has relatively low probabilities for all classes. This either means that the segment does not belong to any of the classes investigated, or it may mean that the segment does not contain voxels with similar features for the investigated classes. The black segment (second row) has the highest probability for grade III, but also some for grade II. The green

segment has a high probability for healthy, but also some for grade II. In comparison, the five segment solution (row 8-12) shows a high probability for grade III for the blue segment. The black segment has a high probability for grade II, the green for grade II and III. The red and yellow segments have a high probability for healthy tissue. If compared with the 3 segment solution, the 5 segment solution has more segments with a high probability for one class and is thus to be favored above the 3 segment solution. figure 6.2c provides the clinician with an easily visualized statistical measure, that can be used to assess the clustering. Together with the MR images and the clinical information, the best clustering can be chosen by an expert in the field of radiology. The five segment solution is selected, because of the very high probability of the blue segment for grade III in the five segment solution and the consistency with the MR images.

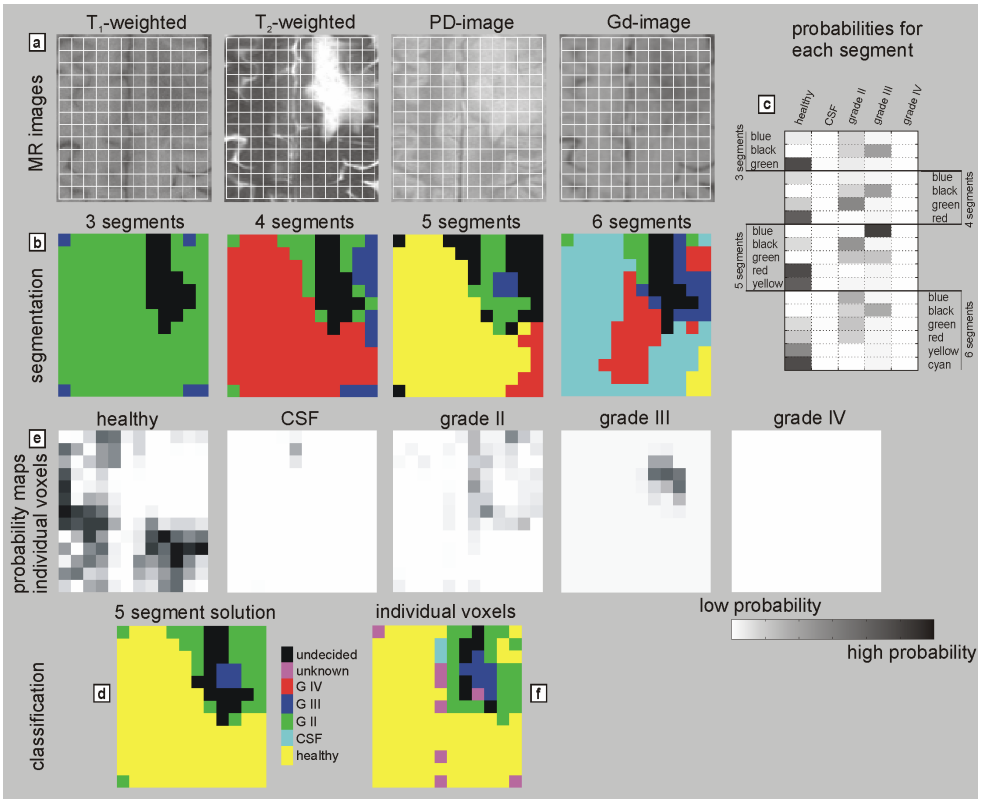


Figure 6.2. Results of the classification strategy for patient I-1229. *The patient has a grade III glial tumor. The I-number corresponds with the INTERPRET database. See the text for the explanation of the figures.*

The classification result of the segmentation is shown in figure 6.2d. From now on, the colors of the segments represent a classification result: black) undecided, purple) unknown, red) grade IV, blue) grade III, green) grade II, cyan) cerebrospinal fluid, and yellow) healthy. Different segments can be classified to the same class, as shown in the example. The former red and yellow segments have been assigned to healthy and are therefore colored yellow. Most of the brain is classified as healthy. In the tumor area, a small grade III classified section (blue) is surrounded by two other sections. The black one is undecided. The green one is classified as grade II.

Next to the probabilities and classification of the segments, also the probabilities and classification of the individual voxels may be of clinical interest. The probability maps of the individual voxels for each class are plotted in figure 6.2e. There are as many maps as there are classes. In each map, the probability of class membership for all voxels for a specific class is plotted in gray. The gray levels are comparable with the levels found for the probabilities of the segments. The interpretation of the maps is very important. If a voxel has a high probability of membership for only one class, then the classification will be reliable. If a particular voxel has a probability for more classes, then the resulting classification will be unreliable. In certain cases it may indicates that the tissue is heterogeneous. Low probabilities for all classes indicates that the underlying tissue type is unknown. Using the classification rules dictated by the classifier, each voxel can be classified. The result of classification is presented in figure 6.2f. The colors represent the same tissue type as in the classification of the segments.

To compare the classification result of the segments and the individual voxels with each other and with the pathology of each patient, a single classification result from each had to be deduced. This single classification result is summarized in table 6.1, and compared with the classification obtained by the radiologist. To obtain a single classification result from the spatially resolved segmented brain, the following rules were applied: a) first, the best segmentation was selected by comparing the MR images, the segmented images and the accompanying probabilities; b) then, the tumor was assigned to the classification of the segment with the highest grade. If the voxels of this segment were not next to each other, but for example, were situated in the corners, then the classification result of this segment was not taken into account. Then, the segment with the next highest grade was validated. For patient I-1229, the five segment result was chosen, thus the tumor was classified as a grade III.

A single classification result of the tumor using the individual voxels was more difficult, because the classification was normally heterogeneous. Classification was performed as followed: a) the number of voxels was counted that were classified the most malignant (with the same grade); b) it was checked whether these voxels were next to each other, and in the vicinity of the tumorous area. If so, the tumor was assigned to this class. If this was not the case, another class was selected based upon the voxels with the second highest grade. This selection method was used, since it often occurred that voxels at the edge were misclassified as grade IV. Hence, to circumvent this artifactal problem, voxels with the same classification had to be next to each other and inside the clearly visible tumorous area.

Table 6.1 (page 129) shows that only one of the cases in the training set is misclassified. I-1232 is classified as a grade II tumor while its pathology is grade III. The independent test case I-1283, is classified as 'unknown', which is correct, since the classifier was not trained to take meningiomas into account. The two cases for which the pathologists could not reach a consensus are also classified to test the system. In both cases, the classification result of segments and individual voxels is in accordance with the morphology of the brain, as seen on the MR images. The healthy tissue and CSF can be clearly identified. The classification result of the tumor of patient I-1281 is grade II for the classification using segments as well as for the individual segments. The case I-1172 is shown in figure 6.3, to illustrate the result of a test case from which no tumor voxels have been selected in the training set.

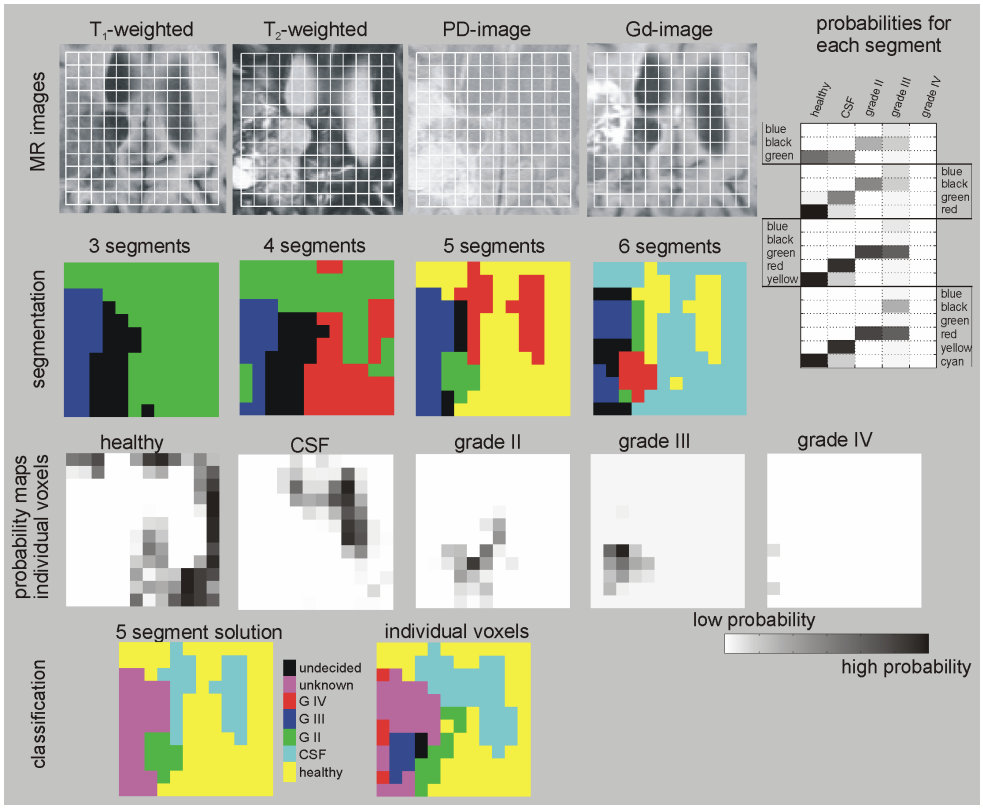


Figure 6.3. Decision support results of patient I-1172.

The MR images for patient I-1172 show a malignant area in the lower left of the brain. This part is isolated in the 3, 4, 5 and 6 segment solution. Only in the 5 and 6 segment solution the ventricles are isolated. Also the probabilities for class membership are higher for the segments in the 5 or 6 segment solution, if compared with the 3 and 4 segment solution. The classification result of the 5 segment solution is plotted on the lower left of figure 6.3. This shows a large 'unknown' segment and a grade II segment. The individual classification shows the heterogeneity of the malignant area. Voxels are classified as grade II, grade III, grade IV and 'unknown'. This reflects the difficulties of the pathologists to validate this patient.

Table 6.1. Classification results using pathology, classification of segmented brain and classification of individual voxels.

I-[a]	Pathology[b]	S[c]	classification segments[d]	classification individual voxels[e]	remarks
1227	II	4	II	II	
1233	II	4	II	II	1 gIV voxel outside tumor
1260	II	4	II	II	3 gIV voxels outside tumor
1275	II	5	II	II	
1278	II	3	II	II	3 gIV voxels outside tumor
1318	II	6	II	II	
1212	III	3	III	III	1 gIV corner voxel
1229	III	5	III	III	
1230	III	3	III	III	4 gIV voxels inside tumor
1232	III	5	II	II	not one gIII voxel or segment; one gIV voxel
1231	IV	4	IV	IV	
1234	IV	5	IV	IV	tumor area is surrounded by 'unknown'
1285	IV	4	IV	IV	
1286	IV	3	IV	IV	all segments and most voxels other than gIV are classified as 'unknown'
1322	IV	5	IV	IV	
1283	MNG	3	'unknown'	'unknown'	
1172	II/III	5	II	IV	tumor area is a mixture
1281	II/III	5	II	II	

[a]: *patient number. The numbers correspond with the INTERPRET database.*

[b]: *Pathology (only the grade if glial) of the tumor as validated by consensus histo-pathologic determination.*

[c]: *The optimal number of segments in the brain as selected by an expert.*

[d]: *See the text for explanation of the single classification deduced from the classification of segments.*

[e]: *See the text for explanation of the single classification deduced from the individually classified voxels.*

The results of the evaluation of the segmentation is schematically represented in table 6.2. For each subject, the number of pathologists is given that selected the 3, 4, 5 or 6 segment result as best. In two cases all radiologists agreed that the same segmentation result corresponded best with their drawing. In one case the radiologists choose between the 4 and 5 segmentation result, and in another between the 3 and 4 segmentation result. Only in one case the radiologists disagreed between 3 segmentation results.

Table 6.2. Results of clinical evaluation

	question 2 of questionnaire[a]				question 3 of questionnaire[b]	
subject	**3 segm.**	**4 segm.**	**5 segm.**	**6 segm.**	**yes**	**no**
I-1229		2	2		4	0
I-1231		4			3	1
I-1260	2	1	1		3	1
I-1278	4				4	0
I-1285	2	2			3	1

[a]: *Question: which segmentation result produced by the clustering corresponds best to the drawing you made of the tumor?*

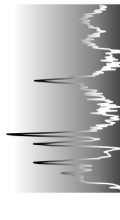
[b]: *Question: does the selected segmentation agree with the drawing you made?*

Remarks of the radiologists were that sometimes the tumor extended outside the STEAM box, causing this part of the tumor not to be taken into account by the segmentation algorithm. Other remarks were that the tumorous part was sometimes divided into multiple segments by the classifier, or that the healthy region consisted of multiple segments. This was considered to be a disadvantage by some, but not all by radiologists. However, usually, multiple segments were combined after the classification procedure because the segments were identically classified. During the clinical evaluation of the clustering this was not taken into account.

In the example case presented in figure 6.2 (I-1229), all radiologists draw an equally large region at the same location. Subsequently, two radiologists draw a smaller region within the first one. These two radiologists favored the five segment result, while the other favored the four segment result.

Discussion

A medical decision support system has been developed to aid clinicians in the diagnosis of patients with a brain tumor. The system condenses spectral information that may be difficult to assess, combines it with MR imaging information and provides a classification result. The classification results are displayed as images, which will facilitate the interpretation and provides information of the heterogeneity of the tumor. Tumors are known to be heterogeneous. Several groups[16, 13-15] have reported that MRSI can play a role in the detection of metabolic changes in the tumor tissue as well as in the surrounding tissue, but mostly univariate methods are used. Only few reports have appeared which discuss the use of multivariate approaches on MRSI data[4, 5, 12] in the classification of brain tumors, whereas the data is highly suitable for multivariate analysis. One of the major advantages of MRS is that it allows the observation of a large number of metabolites simultaneously. During multivariate analysis this information is combined (possibly including MRI or other information) and pattern recognition[8, 10, 27] is used to identify and interpret the meaningful regularities in the data. Pattern recognition has shown to provide good classification results on spectroscopy data[6, 11, 28] and MRI data[29], and the combined findings of MRI and single voxel spectroscopic data by experts in the diagnosis of intracranial mass lesions has shown to increase the correct diagnosis[30]. Therefore, the combination of MRI and MRSI data in a multivariate classification approach may serve as a solid basis for reliable automated brain tumor classification. Since our results show a correct classification of 14 of the 15 cases, this indeed seems to be the case. Especially, since only a maximum of 30% of the voxels of each patient were within the training set.



One of the difficulties in the use of both spectral and image data in tumor classification, is the harmonization of both data sources for input. Since the images only add four extra variables, feature reduction applied to exclusively the spectra seems a meaningful solution. In our approach a simple quantitation of relevant metabolite peaks was used. This, to incorporate prior biochemical knowledge into the feature reduction method. Next to feature reduction on the spectroscopic data, the spatial resolution of the MRI and MRSI have to be matched. Since the two are acquired with a large difference in spatial resolution the selection has to be chosen carefully. Lowering the MRI resolution will lead to loss of information, but when the MRI resolution

is kept, the registration of the images needs to be very accurate in order to avoid mixing of pixels (due to patient movement). At a higher resolution, the weighting of the image variables is larger than the MRSI variables. Details in a region within a spectroscopic voxel would be due to the image variables alone. Furthermore, when increasing the resolution of the MRSI data, the information from each MRSI pixel is still acquired from a large volume. This will lead to pixels which are described by variables which contain information from a much larger volume than the pixel volume itself. Since normally the tumor covers multiple MRSI voxels, it was decided to use a resolution of twice the true MRSI resolution. In this way we used the spectral variables optimally, avoiding problems with co-registration of MRI images, modeling complexity and computation time.

To further increase the reliability of the system, it is important to calculate probability values for the classification. High probabilities for a tumor type in a specific region may help to localize the most malignant part of the tumor, and thus play a role in the assessment of heterogeneity. This is especially helpful in guiding the clinician towards the best location to perform a biopsy.

A consensus diagnosis can be considered as another method to improve the classification result[9]. Consensus diagnosis involves the analysis of the same data by several multivariate methods. In our approach the classification of the segments and individual voxels is performed by the same method. However, the segmentation algorithm is totally independent to the classifier. We have found in several cases that the probability of classification of a segment in the tumor was higher than any of the probabilities of the individual voxels. Since the tumor normally originates from one type of cell and is often larger than one voxel, it could be advantageous to combine voxels within the tumor during the classification process. However, instead of averaging all voxels in the tumor manually, the segmentation algorithm only averages those voxels which share features. Therefore, it reckons with the spatial heterogeneity of the tumor. The algorithm needs no prior knowledge, nor normalization between patients. It is therefore independent of a training set and will always provide relevant information of the location and different fractions within the tumor, even in the case if classification is not possible. Evaluation of the segmentation is very difficult as there is no golden standard available. However, most of the time the ventricles and the contra-lateral healthy brain were found as separate segments. Often the tumor area itself was divided in several segments, which can be explained by the presence of heterogeneous

tissue. Radiologists who were asked to validate the segmentation were positive on the results. For all five investigated cases, the radiologists were able to point out a clustering result that to their opinion covered the region that they had drawn as malignant.

Classification of the segments is convincing for all patients. Always at least one segment is found which covers most of the tumor and that is classified in agreement to the pathology of the patient, except for patient I-1232. This patient is misclassified as a grade II, while it has been validated as a grade III. Calculation of an objective criterion for the evaluation of each segment is difficult. However, a comparison of the classified segments with the high resolution images and the pathology seems to indicate that classification of the segments is robust. In the case of patient I-1172, a consensus regarding the pathology was not reached by the three radiologists who validated the patient. The radiologists opinions disagreed between tumor type and grade: a) astrocytoma grade III, b) astrocytoma grade II, and c) oligodendroglioma grade II. The segmentation of this patient only seemed meaningful in the five and six segment solution and exhibited multiple segments in the malignant area. The probabilities of class membership for the blue and black segment were low for all classes in the five segment solution. These segments could not be classified ('unknown'). The relatively small green segment had a high probability of class membership for the grade II as well as grade III class, but was classified as grade II. The classification of the individuals voxels showed grade II, III, IV and 'unknown' in the malignant area. The classification reflected the difficulties of the radiologists in typing the malignancy of this patient.

The classification results of the individual voxels was for each patient in agreement with the pathology of the tumor, except again for I-1232. However, in some cases voxels in the vicinity of the tumor or at the border of the STEAM box were misclassified as grade IV, which could lead to misclassifications. This shows the importance of the classification of the segments, which did not show this kind of errors. The consensus of both methods seems to improve the reliability. The probability values of the individual voxels are also important. They provide information about the reliability of classification and sometimes the heterogeneity of the tissue. If a voxel has a high probability for a specific tissue type and only for this tissue type, then the tissue at that location will be relatively homogeneous. If the probabilities of a particular voxel are the same for two tissue types, then the

reliability of classification will be lower. If the probabilities of class membership are the same for two tissue types which both can be present in the brain, then this can also indicate tumor heterogeneity. If the probabilities are high for two totally different tumor types, then the reliability of classification is low. This phenomenon is also reflected in the probability maps of patient I-1172 (figure 6.3). The healthy and CSF regions have high probabilities for healthy and CSF, which indicates homogeneity of the tissue in that area. In the right part of the brain, several voxels have high probabilities for grade II as well as grade III, indicating tissue heterogeneity.

Acknowledgment

The authors acknowledge J.H.C.L. Hendriks (MD), A.R. de Vries (MD) and A.M.A. Boerboom (PhD) from the University Medical Center Nijmegen for clinical evaluation. The EU funded INTERPRET project (IST-1999-10310; http://carbon.uab.es/INTERPRET/) is gratefully acknowledged.

References

[1] Preul MC. Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy. Nature Medicine 1996; **2**(3): 323-325.

[2] Nelson SJ. Analysis of Volume MRI and MR Spectroscopic Imaging Data for the Evaluation of Patients With Brain Tumors. Magnetic resonance in medicine 2001; **46**(2): 228-239.

[3] Meyerand ME, Pipas JM, Mamourian A, Tosteson TD, Dunn JF. Classification of biopsy-confirmed brain tumors using single-voxel MR spectroscopy. American Journal Neuroradiology 1999; **20**: 117-123.

[4] Szabo De Edelenyi F, Rubin C, Estève F, et al. A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors:nosologic images. Nature Medicine 2000; **6**(11): 1287-1289.

[5] Preul MC, Caramanos Z, Leblanc R, Villemure JG, Arnold DL. Using pattern analysis of in vivo proton MRSI data to improve the diagnosis and surgical management of patients with brain tumors. NMR in biomedicine 1998; **11**: 192-200.

[6] Tate AR, Majós C, Moreno A, Howe FA, Griffiths JR, Arús C. Automated classification of short echo time in in vivo 1H brain spectra: a multicenter study. Magnetic Resonance in Medicine 2003; **49**: 29-36.

[7] Wehrens R, Simonetti AW, Buydens LMC. Mixture modelling of medical magnetic resonance data. Journal of chemometrics 2002; **16**(6): 274-282.

[8] Hagberg G, Burlina AP, Mader I, Roser W, Radue EW, Seelig J. In vivo proton MR spectroscopy of human gliomas: definition of metabolic coordinates for multidimensional classification. Magnetic Resonance in Medicine 1995; **34**: 242-252.

[9] Somorjai RL, Nikulin AE, Pizzi N, et al. computerized consensus diagnosis: A classification strategy for the robust analysis of MR spectra. I. Application to 1H spectra of thyroid neoplasms. Magnetic Resonance in Medicine 1995; **33**: 257-263.

[10] El-Deredy W. Pattern recognition approaches in biomedical and clinical magnetic resonance spectroscopy: a review. NMR in Biomedicine 1997; **10**: 99-124.

[11] Herminghaus S, Dierks T, Pilatus U, et al. Determination of histopathological tumor grade in neuroepithelial brain tumors bu using spectral pattern analysis of in vivo spectroscopic data. Journal of Neurosurgery 2003; **98**: 74-81.

[12] Simonetti AW, Melssen WJ, van der Graaf M, Postma GJ, Heerschap A, Buydens LMC. A Chemometric Approach for Brain Tumor Classification Using Magnetic Resonance Imaging and Spectroscopy. Analytical Chemistry 2003; **75**(20): 5352-5361.

[13] Furuya S, Naruse S, Ide M, et al. Evaluation of metabolic heterogeneity in brain tumors using 1H-chemical shift imaging method. NMR in biomedicine 1997; **10**: 25-30.

[14] Segebarth C, Balériaux D, Luyten PR, den Hollander JA. Detection of metabolic heterogeneity of human intracranial tumors in vivo by ^{1}H NMR spectroscopic imaging. Magnetic Resonance in Medicine 1990; **13**: 62-76.

[15] Fulham MJ, Bizzo A, Dietz MJ, et al. Mapping of brain tumor metabolites with proton MR spectroscopic imaging: clinical relevance. Radiology 1992; **185**: 675-686.

[16] Mader I, Roser W, Hagberg G, et al. Proton chemical shift imaging, metabolic maps, and single voxel spectroscopy of glial brain tumors. MAGMA 1996; **4**: 139-150.

[17] Nelson S, Huhn S, Vigneron DB, et al. Volume MRI an MRSI techniques for the quantitation of treatment response in brain tumors: presentation of a detailed case study. Journal of Magnetic Resonance Imaging 1997; **7**: 1146-1152.

[18] Minguillon J, Tate AR, Arus C, Griffiths JR. Classifier Combination for *in vivo* Magnetic Resonance Spectra of Brain Tumours. Lecture notes in computer science 2002; **2364**: 282-292.

[19] International network for pattern recognition of tumours using magnetic resonance: http://carbon.uab.es/INTERPRET/interpret_home.shtml

[20] Klose U. In vivo proton spectroscopy in presence of eddy currents. Magnetic Resonance in Medicine 1990; **14**: 26-30.

[21] Simonetti AW, Melssen WJ, van der Graaf M, Heerschap A, Buydens LMC. Automated correction of unwanted phase jumps in reference signals which corrupt MRSI spectra after eddy current correction. Journal of magnetic resonance 2002; **159**: 151-157.

[22] Pijnappel WWF, van den Boogaart A, de Beer R, van Ormondt D. SVD-based quantitation of magnetic resonance signals. Journal of magnetic resonance 1992; **97**: 122-134.

[23] Campbell ID, Dobson CM, Williams R.J.P., Xavier A.V., Resolution enhancement of protein MR spectra using the difference between a broadened and a normal spectrum, Journal of Magnetic Resonance 11: 172-181, 1973

[24] McLachlan GJ, Peel D (2000). *Finite Mixture Models*, Wiley: New York.

[25] Schwarz G. Estimating the dimension of a model. Annals of statistics 1978; 6: 461-464

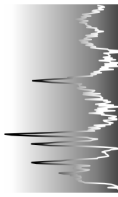
[26] Mahalanobis PC. On the generalized distance in statistics, Proc. Nat. Inst. Sci. (India) 12 (1936) 49-55

[27] Witjes H, Simonetti AW, Buydens LMC. Better brain imaging with chemometrics. Analytical chemistry 2001; **73**(19): 548A-556A.

[28] Somorjai RL, Dolenko B, Nukulin A, et al. Distinguishing normal from rejecting renal allografts: application of a three-stage classification strategy to MR and IR spectra of urine. Vibrational Spectroscopy 2002; **28**: 97-102.

[29] Reddick WE, Glass JO, Cook EN, David Elkin T, Deaton RJ. Automated Segmentation and Classification of Multispectral Magnetic Resonance Images of Brain Using Artificial Neural Networks. IEEE transactions on medical imaging 1997; **16**(6): 911-918.

[30] Möller-Hartmann W, Herminghaus S, Krings T, et al. Clinical application of proton magnetic resonance spectroscopy in the diagnosis of intracranial mass lesions. Neuroradiology 2002; **44**: 371-381.



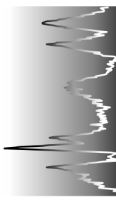
Chapter 7

Feature reduction and combination of MR spectroscopic and imaging data for improved brain tumor classification

AW Simonetti[♣], *WJ Melssen*[♣], *F Szabo de Edelenyi*[♣], *JJA van Asten*[♦], *A Heerschap*[♦], *and LMC Buydens*[♣]

[♣] **Laboratory for Analytical Chemistry, University of Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands**

[♦] **Department of Radiology, University Medical Center Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands**



Abstract

The purpose of this paper is to investigate feature selection on several sources of MR data, to provide an optimal basis for brain tumor classification. Suppressed and unsuppressed short echo time magnetic resonance spectroscopic imaging data (MRSI) and MR imaging data from 28 subjects were used during the investigations. Four different feature reduction procedures were applied on the MRSI data; simple quantitation, principal component analysis, independent component analysis and LCModel. From the unsuppressed MRSI data, the intensity of the water signals were calculated. Different combinations of the MRSI features, the water signal and the MRI features were used as input for a classification procedure, that was setup to classify four types of tumor, healthy brain tissue, and cerebrospinal fluid. The classification results were compared to investigate the best feature selection procedure.

Conclusions are that feature reduction on the MRSI data, if carefully executed, is not critical for the classification result. However, combination of features from MRSI and MRI data improves the classification result with approximately 15%. The addition of features from the water signal additionally increases the classification result, although not significantly. The combination of data from different sources (including for example clinical information) may be very important for the result of brain tumor classification, particularly if a large number of tumors is needed to be classified simultaneously.

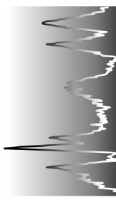
Keywords: magnetic resonance spectroscopy, data processing, feature selection, classification, pattern recognition, brain tumor.

Introduction

The correct diagnosis of tumor type and grade is very important for the treatment of patients with a brain tumor. At the moment, the diagnosis is mostly made through a histopathologic determination of a tissue sample taken from a biopsy. Imaging techniques, like magnetic resonance imaging (MRI), are used to locate the position and extent of the tumor and are a further help in the diagnosis.

To diminish the use of an invasive biopsy, which is very stressful for the patient, the usefulness of proton magnetic resonance spectroscopy (MRS) for tumor typing is under investigation[1-3]. Proton MRS provides completely different information as compared to imaging tools, but can be obtained with most 1.5 T MRI instruments used for routine radiology. The acquired MR spectra provide localized metabolic information related to cell membrane proliferation, neuronal damage and energy metabolism of brain or tumor tissue. The spectra contain information from important brain metabolites like glutamate, myo-inositol, choline, creatine, N-acetyl aspartate (NAA), lactate and fatty acids[4]. Proton MRS can be extended to an imaging mode called proton MRS imaging (MRSI), in which the acquisition of multiple voxels provides spatially encoded chemical information from the brain. MRSI, in combination with MRI[5-9], could possibly be used to resolve tumor location, type and grade in an automated noninvasive way. However, the large amount of data acquired during the investigation of the patient, and the difficult interpretation of the spectroscopic data requires the development of classification strategies that are based on pattern recognition.

Four stages can be defined in the strategy of developing automated tumor classification models with the use of pattern recognition. The first comprises of a selection of training samples (training set) for the tumor classes to be classified. Optimally, for each class also an independent test set is selected, in order to test the accuracy of the classifier. The second step consists of preprocessing on the data. This may involve several steps of which the following are the most common; a) Eddy current correction[10], b) water filtering with HLSVD[11], c) adjustment of begin time (first order phase correction in time domain), d) Fourier transformation, and e) normalization with a separately obtained water signal.



The third step involves feature selection of the highly dimensional spectroscopic data. During feature selection, variables are selected from the data, which are meaningful for the classification. This is especially important when spectroscopic data is used in the classification, since in this case the number of variables normally exceeds the number of objects. Feature selection should only retain the information that optimally discriminates the tumor classes under investigation. Unfortunately this is not a straightforward task. Feature selection also implies that features from different sources of data are combined. In the case of data obtained from clinical MR systems, this is not only reserved to the combination of water suppressed proton MRSI and MR imaging data. Also the water signal which is obtained from the unsuppressed MRSI can be used, or spectroscopic information from other MR sensitive nuclei, like ^{31}P.

The last step in automated tumor classification, involves the actual classifier. If the feature selection has been sufficient, the shortest distance to the centroids of the classes, or the k-nearest neighbors[12] can be used as classifier. The data can also be used as input for methods which further reduce the dimensionality by finding optimal separable directions, e.g. linear discriminant analysis or quadratic discriminant analysis. There are much other, more specialized methods, for example based on density functions[13], neural networks[14] or support vector machines[15].

In the search for the best strategy for brain tumor classification, feature selection has not received much attention. In this paper, several feature selection methods are compared. First, features are selected from MRSI data with four different feature reduction methods. Also features are selected from MRI data and from water signals obtained from unsuppressed MRSI data. Then, combinations of the selected features are used to build classification models. Last, the classification results of an independent test set are compared to find the best feature selection method.

Experimental

Data acquisition protocol. This study has been approved by the ethical committee of the university medical center and from all patients written informed consent was obtained. Data of 24 patients with a brain tumor and 4 volunteers were used in the study. Each patient case had passed strict quality control and validation procedures, including consensus histopathologic

determination. 21 patients were diagnosed with a glial brain tumor of specific grade and 3 with a meningioma.

MR data from each patient was acquired following a distinct defined acquisition protocol setup by the INTERPRET consortium[16]. All measurement were performed on a 1.5 T Siemens Vision whole body system, using a CP-head coil. The protocol consisted of the acquisition of conventional T_1-weighted, T_2-weighted and proton density weighted images (PD image). This was followed by a T_1-weighted image after intravenous administration of Gadolinium-DTPA (Gd image), and concluded by water suppressed and unsuppressed proton MR Spectroscopic Imaging (^{1}H-MRSI).

Transversal MR images were collected using the following parameters: 23 slices of 5-mm thickness, inter-slice distance 1.5 mm, FOV 230 mm, matrix 256x256. For the T_1-weighted images (TE/TR=15/644 ms) a standard spin echo sequence was used. The PD image (TE/TR=16/3100 ms) and the T_2-weighted images (TE/TR=98/3100ms) were simultaneously acquired using a turbo spin echo sequence. After a bolus injection of 15 ml 0.5 M Gd-DTPA a T_1-weighted measurement was performed to obtain the Gd image. Eventually, only the four images (with different contrasts) acquired from the same location as the MRSI were retained. Images just below or above were discarded, since they were not totally within the MRSI slice.

The MRSI were acquired using a 2D STEAM ^{1}H-MRSI sequence with short echo time. The STEAM box was positioned in a transversal plane through the brain showing the largest Gadolinium-DTPA enhanced tumor area in the Gd image. It was placed entirely in the brain parenchyma avoiding leakage of disturbing signals from fatty tissue surrounding the skull. MRSI parameters were: 16 x 16 x 1024 samples, TR/TE/TM = 2000 or 2500 ms/ 20 ms / 30 ms, slice thickness = 12.5 or 15 mm, FOV = 200 mm, spectral width = 1000 Hz and NS = 2 (1 for the MRSI without water suppression).

Data preprocessing. The MRI and MRSI data were processed in the same unique way for each case and automated where possible. The preprocessing procedures are schematically presented in figure 7.1a. For the suppressed (metabolite) MRSI data (left column of figure 7.1a); K-space data was filtered by a Hanning filter of 50% using the LUISE software package (Siemens, Erlangen, Germany), followed by zero filling to 32 X 32 and 2D Fourier transformation to obtain time domain signals for each voxel. Then, each voxel within the STEAM box was corrected for eddy current effects in the spectra using a method which prevents the occasional occurrence of eddy current

correction induced artifacts[17]. This was followed by HLSVD filtering[11] to remove the residual water signal between 4.3 and 5.5 ppm. Next, a simple but efficient baseline correction with a filter width of 5 ms was applied, that removed broad resonances[18]. To correct for first order phase effects (in principle the zero order phase is removed by the eddy current correction), the first order phase of the mean spectrum as calculated from all spectra within the STEAM box of each patients MRSI dataset, was manually optimized. This correction value was then applied to each separate time domain signal of that dataset. The resulting data consisted of the preprocessed free induction decays (FIDs) for each voxel within the STEAM box of the patient.

In order to combine information from the spectroscopic data with information from the MR images they need to have the same resolution and have to be aligned. The MR images were aligned with respect to each other, by successively shifting the T_1- weighted and Gd image with respect to the PD image until a maximum of spatial correlation was reached. The MRSI spectroscopic grid was assumed to be aligned with the PD image, since they were acquired consecutively. After alignment of the MR images, the image pixels which did not fit within the boundary of the STEAM box were discarded. Then, the resolution of the remaining part of the MR images was lowered to the resolution of the MRSI grid. This was performed by averaging the image pixels which were covered by each spectroscopic voxel.

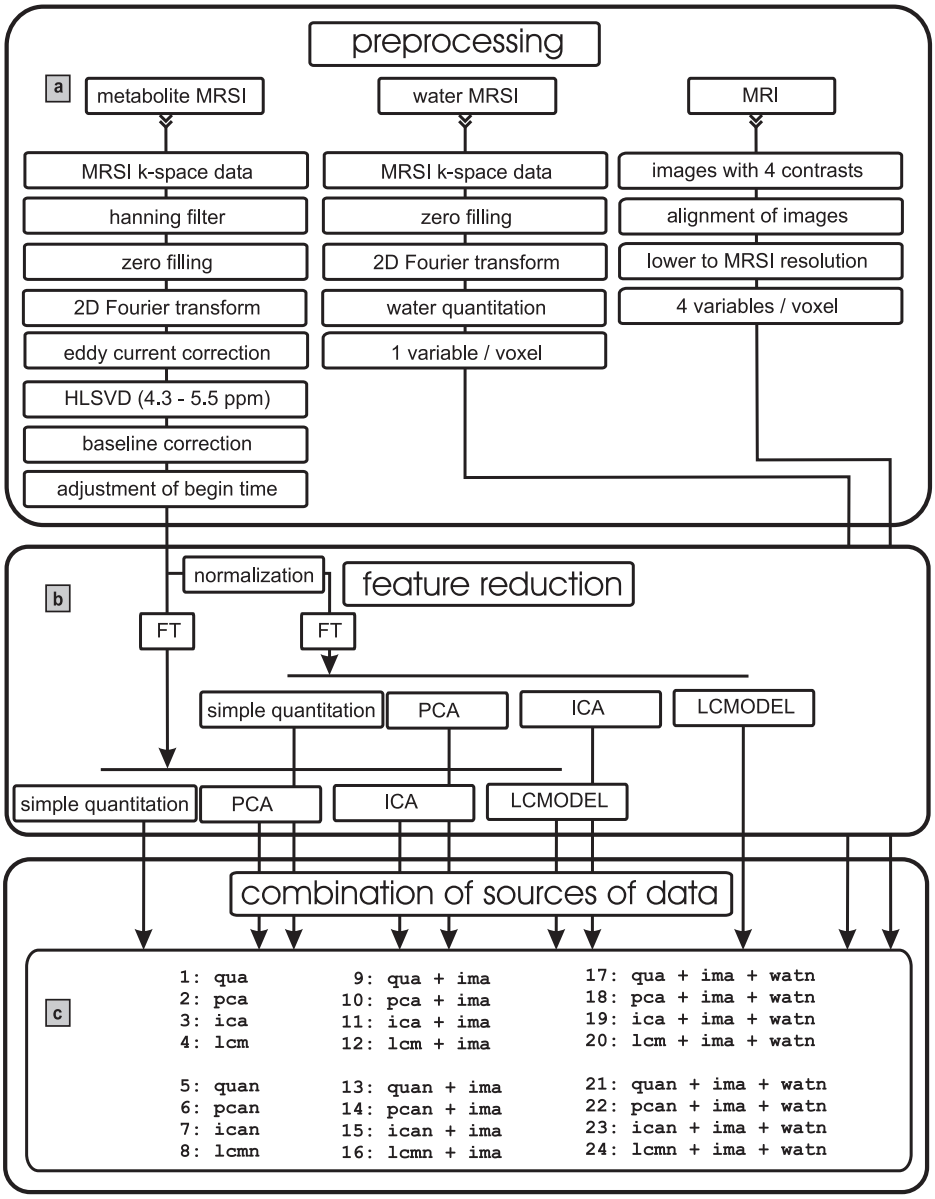


Figure 7.1 Overview of data preprocessing and feature selection. *A schematic overview of the preprocessing of the metabolite MRSI, water MRSI and MRI is shown in figure a. The feature reduction methods used on the metabolite MRSI are shown in figure b. The feature reduction is performed on the non-normalized data as well as on the water normalized data. Classification results were calculated for each of the shown combinations of different sources of data in figure c.*

The unsuppressed water signal. The unsuppressed (water) MRSI was used for 3 specific reasons. First during preprocessing, to perform eddy current correction on each voxel. Secondly, the water signal for each voxel was quantitated. This was done as follows: K-space data were filtered by the same Hanning filter as the metabolite MRSI data, followed by zero filling to 32 X 32 and 2D Fourier transformation to obtain time domain signals for each voxel. Then, the amplitude of the water signal was estimated with software provided by dr. L. Vanhamme (Katholieke Universiteit Leuven, Department of Electrical Engineering, SCD-SISTA, Kasteelpark Arenberg 10, 3001 Heverlee, Belgium). This software models the water signal in the time domain by a Voigt model, with an additional first order term that corrects for eddy currents. HSVD is used to obtain initial estimates of the model parameters, that are subsequently used as starting values in a nonlinear least squares optimization algorithm to obtain the final parameter estimates. Third and last, the mean water signal of all voxels within the STEAM box of each subject was calculated. This value was used as an inter patient normalization factor. Water normalization of the metabolite MRSI was performed by the division of each metabolite FID in a patients data set by its calculated mean water signal. Also the calculated water signal of each voxel in a patients data set, was divided by the mean water signal, to obtain inter patient normalized water signals for each voxel.

construction of the data set. To create a sufficiently large data set, several MRSI voxels from each subject were selected and assigned to one of the classes under investigation. The glial tumors were divided into three classes, according to the rules of the World Health Organization (WHO): grade II (10 cases), grade III (4 cases) and grade IV (7 cases) tumors. Furthermore, classes for meningioma (MNG, 3 cases), healthy brain tissue, and cerebrospinal fluid (CSF) were constructed. The data for healthy tissue were selected from the contra-lateral brain region of 4 patients and from 4 healthy volunteers. Data for the CSF were selected from 8 patients. Only CSF voxels that were not in close contact with the tumor region were selected. This led to a data set containing 6 classes and data from 669 voxels.

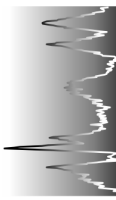
All voxels in the data set were labeled, so when a specific feature selection method was applied, the data could be split afterwards into a training- and test-set, which for each feature selection method consisted of the same voxels. In table 7.1 an overview is presented of the total number of selected voxels for the different classes and the number of voxels in the training- and test-set, respectively.

Table 7.1. Selected voxels for each class in the data set.

tumor class	data set	training set	test set
healthy	218	153	65
CSF	100	70	30
grade II	176	123	53
grade III	57	40	17
grade IV	70	49	21
MNG	48	34	14
total	669	469	200

Investigated feature reduction methods for the MRSI data. After preprocessing, the information from each voxel in the data set consisted of a metabolite FID of 1024 complex values, four variables from the MR images and one variable from the water signal. To use this information in a classification approach, at least feature reduction had to be performed on the high dimensional metabolite FID. In figure 7.1b an overview is given of the feature reduction methods used to extract new features from the metabolite MRSI data. Feature reduction was performed on the non-normalized and water normalized FIDs of the metabolite MRSI data. The feature reduction methods are explained in more detail below.

After feature reduction, several combinations of the data were made, and used as input for the classifier. The output of the different feature reduction approaches were coded in the following way; 'qua' for simple quantitation, 'pca' for principal components analysis, 'ica' for independent component analysis, and 'lcm' for quantitation using LCModel. The feature reduction results of the water normalized metabolite MRSI data were coded 'quan', 'pcan', 'ican', and 'lcmn'. The data containing the image information were coded 'ima'. The data containing the inter patient normalized water signals were coded 'watn'. Figure 7.1c provides an overview of all 24 used combinations.



a) Simple quantitation. Each FID in the data set was Fourier transformed to obtain a spectrum from which the real part of the region between 0.5 to 4.0 ppm was retained (230 spectral points). An example of this part of the spectrum of 230 points is shown in figure 7.2. Then intensities of eight spectral ranges encompassing signals of important brain metabolites were determined by integrating the intensities within a window of 0.13 ppm (as indicated by the gray bands in the spectrum of figure 7.2). The integration regions were chosen around the expected chemical shift of the main resonances of the following metabolites; myo-inositol (3.56 ppm), glucose (3.44 ppm), choline (3.20 ppm), creatine (3.02 ppm), glutamate/glutamine (2.20 ppm), N-acetyl aspartate (NAA, 2.02 ppm), lactate/fatty acids (1.33 ppm) and fatty acids (0.90 ppm).

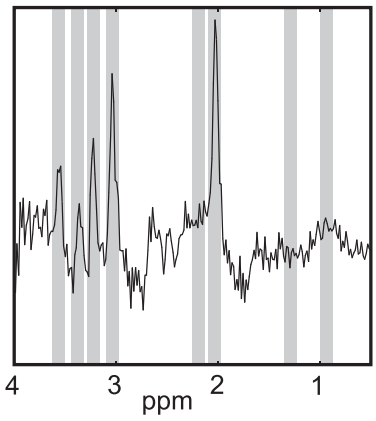


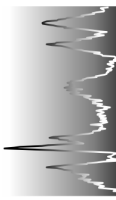
Figure 7.2. Example of spectrum and quantitated regions

b) PCA. For the principal component analysis (PCA), also the real part of the spectral region between 0.5 to 4.0 ppm was retained (230 spectral points). All 230 points were used as input for the PCA. No extra scaling was performed. PCA finds directions in the data that explain most of the variance. These new axes are linear combinations of the original ones, are orthogonal and ordered with respect to the amount of explained variance[12]. After PCA, the first eight principal components, that explained 82.05 % of the variance were retained. The number of principal components to use is normally not straightforward. We used eight to make a comparison with the eight variables obtained by the quantitation of important spectral ranges as described above.

c) ICA. We also used independent component analysis (ICA)[19] as a feature reduction technique to select 8 variables from the original 230 spectral points. During ICA, components that are linear combinations of the original variables are obtained that have the property to be as independent as possible. Theoretically, this means that all information within one component is from one independent source, and that another component cannot contain information from that source. Several algorithms have been proposed in the literature to perform ICA. In this study, the FastICA algorithm[20] is used, because it is fast, relatively robust and easy to use.

First, the dimensionality of the data was further reduced by PCA to 40 components (89,25 % of eigenvalues) prior to the application of ICA, as suggested by A. Hyvärinen[21]. Then, 8 independent components were calculated using a symmetric approach and a skew nonlinearity function. We chose to calculate 8 components to match the number of metabolites quantitated and therefore facilitate the comparison of the different feature selection techniques. However, some additional study has to be performed to know if this number of components is optimal for the classification process. The algorithm allows to obtain the independent components as well as the mixing matrix containing the quantities of each component present in each voxel. The 8 scores of each voxel were then used as new variables.

d) LCModel. The MR spectra in the data set were also analyzed with the LCModel 5.2-2 software package[22]. LCModel uses a linear combination of *in vitro* metabolite model spectra to fit the *in vivo* measured spectra. The *in vitro* metabolite spectra were separately obtained under similar experimental conditions as the *in vivo* measurements. Input of the LCModel was the time domain data of each metabolite FID in the data set, and the *in vitro* linear combination of metabolite model spectra, also referred to as 'basis set'. The LCModel[23] program combines constrained least squares analysis with constrained regulation to optimize the estimated metabolite parameters with a line shape convolution to correct for *in vivo* distortions and a baseline correction of Cubic B-splines with equally spaced knots. Zero and first order phase correction and reference shift correction are included in the model.



In this case, the zero order phase correction was fixed and the 'basis set' was extended by two frequency shifted singlets to simulate the lipid resonances at 0.90 and 1.33 ppm. This enabled additional parameter estimates of these lipid signals. The metabolite concentrations of NAA+NAAG, Cho+GPC+PCh, Cre, Lac+Lip(1.33), Glc, Ins, Glu+Gln and Lip(0.90) were

retained, to provide 8 variables for each FID that reflected the variables obtained by the simple quantitation. Since the classification model was not able to deal with 'missing values', always all concentration values were used, even when the standard deviation of a metabolic compound was higher than 25%. Figure 7.3 shows an example of a fit obtained by LCModel.

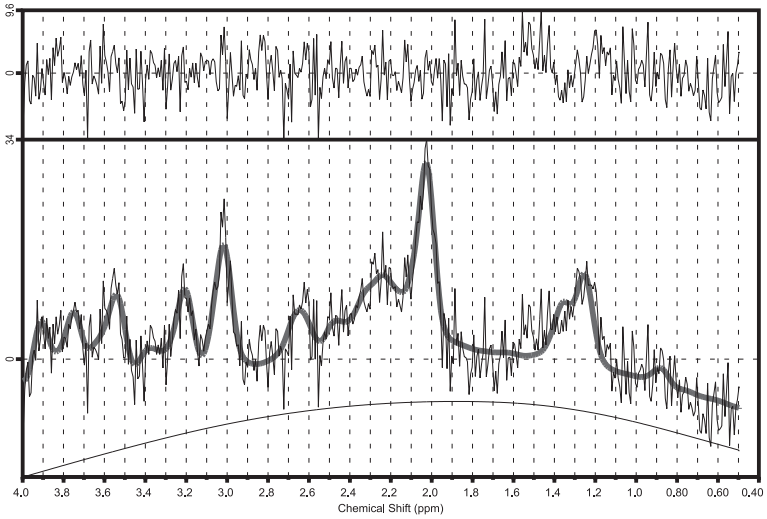


Figure 7.3. Example of LCModel result on a spectrum.

Classification. A classification method has been used[5] that was specifically developed to classify glial tumors, healthy tissue and CSF. This method generates plots that give an impression of the distance of each class with respect to all other classes in the investigated feature space. These so-called distribution plots are important in the assessment of the input data. To be able to calculate the classification results, the classifier was used in its simplest form, which classified a new object to the class it had the closest mahalanobis distance to.

One data set was classified with the use of classification boundaries. These boundaries defined the borders of the classes in feature space, and were setup by setting critical distance values for each class. The critical distances were obtained by inspection of the distribution plots. The critical distance for a specific class was that distance that included most of the objects of the class, while excluding the maximum number of objects of all other classes. During classification, if an object was, for example, closest to the centroid of meningioma, but exceeded the critical distance, then it was classified as

'unknown'. For the grade II and III class, an extra distance rule was setup, since they were found to show considerable overlap. If a new object was close to the grade II centroid, but close to the grade III centroid as well, then it was classified as 'undecided' between grade II and III.

The classifier has been trained with the training set for each different combination of input data. The classification result of the test set was compared with the result obtained by pathology. A classification table was constructed from which the α-errors of each class and a mean α-error of all classes was calculated. The same was done for the β-errors of each class and the mean β-error of all classes. The α-error is a measure of the number of objects of a specific class that are falsely assigned to other classes. The β-error is a measure of the number of objects falsely assigned to that specific class.

The resulting distribution plots and classification tables were compared with each other to estimate the best feature selection procedure. The data from the best feature selection procedure was also used as input for the classifier containing critical distances. This was done to investigate whether the misclassified objects could be successfully identified with the use of selection rules.

Results

The water signal and the effect of normalization. To get an impression of the water signal distribution, as acquired with short echo time MRSI, the water signal image of a patient with a grade IV tumor is compared with its PD image, and the choline, creatine and NAA metabolic images obtained from the metabolite MRSI (figure 7.4). The metabolic images are obtained using the simple quantitation method on the preprocessed spectra. There is clearly structure in the water signal image. The tumorous region exhibits higher water signal (white areas, lower right) than the healthy white/gray matter region. The water signal in the CSF region is lower than in the healthy white/gray matter region. Also an overall lower water signal is visible at the border of the water image. This phenomenon is also visible in the metabolic images. However, the contrast in the water signal image is not as high as the contrast in the metabolic images, indicating that differences in water signal are not so high in the water signal image.

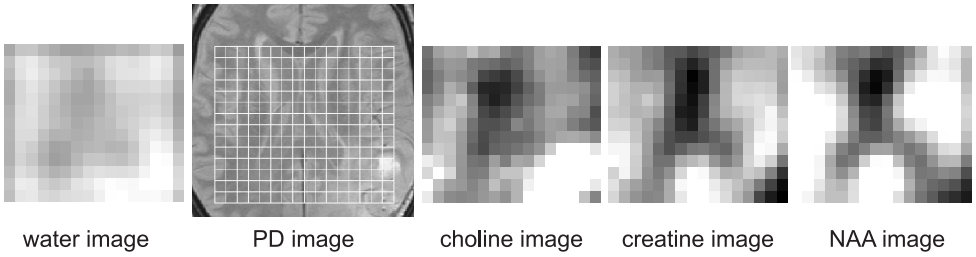


Figure 7.4. Comparison of the water signal image, the proton density image and three metabolic images of a patient with a grade IV tumor.

To estimate whether the water signal had discriminating power, box plots were constructed that provide an overview of the water signal in the voxels selected for the six classes in the data set (figure 7.5). An increase of water signal is clearly visible between the voxels from healthy tissue and tumorous tissue. The water signal of CSF voxels is lower than the water signal of voxels from healthy tissue. Within the glial tumors, there is a gradual increase of water signal when the grade of the tumor increases. The water signal of the meningioma voxels is elevated with respect of healthy tissue, and comparable with the water signal from the grade II glial tumors.

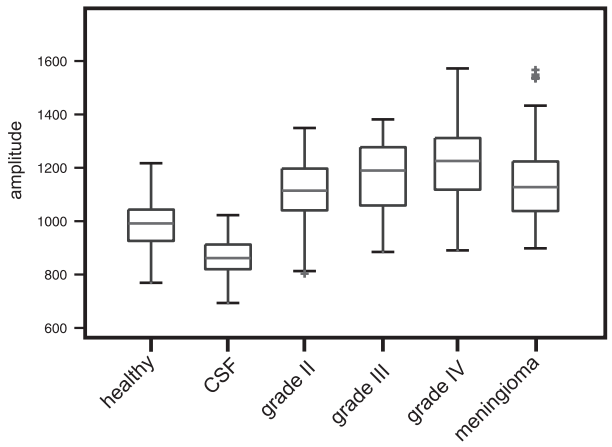


Figure 7.5. Box plots show the water signal levels in the voxels of the six classes of the data set. *The median value for each class is shown by the middle horizontal bar. The box around each median indicate the upper and lower quartiles of the data. The whisker extends to the most extreme data points within 1.5 times the interquartile range of the box. The + signs are data points outside the whiskers.*

To see the effect of inter patient normalization, the non-normalized values of each variable were compared with its normalized values. strikingly, the standard deviation of the values within each class for a certain variable did not show a significant change, when it concerned variables obtained from the metabolite MRSI data (e.g. from simple quantitation, pca, ica and LCModel). Also the means did not change significantly.

However, there were differences visible if the normalization of the water signals was performed. In the example, the non-normalized water signals of all the voxels in each class are plotted (figure 7.6a). Figure 7.6b shows the normalized values. In the normalized case, the values within each class have a smaller distribution. The standard deviation of the values in the healthy class drops from 155 to 92. Also the standard deviations of the grade II, III and IV class drop significantly. The standard deviations for CSF and MNG are in the same order after normalization.

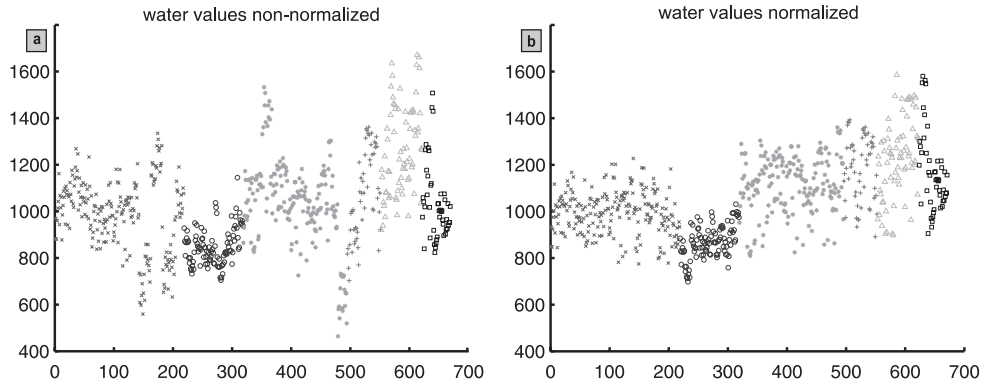


Figure 7.6. Comparison of the non-normalized (a) and normalized (b) water signals of the voxels selected in the data set. *The water signal of the 'healthy' voxels in the data set are labeled with crosses. The CSF with circles, grade II with asterisks, grade III with plusses, grade IV with triangles and MNG with squares. After normalization, the spread within the classes is smaller for the 'healthy', grade II, grade III and grade IV class.*

Since it is not very informative to show all classification tables for all combinations of features, table 7.2 is constructed, that only shows the mean α-error and mean β-error of the independent test set for each feature selection method. The errors are percentages and the values are not corrected for the class sizes. From this table, several feature selection methods were selected which are discussed in the following sections.

Table 7.2. classification errors on the test set using different feature selection methods.

feature selection method	mean α-error (%)	mean β-error (%)
qua	21	12
pca	22	11
ica	26	13
lcm	26	21
quan	22	8
pcan	23	15
ican	27	12
lcmn	27	25
qua + ima	6	4
pca + ima	10	4
ica + ima	11	7
lcm + ima	11	8
quan + ima	6	4
pcan + ima	10	6
ican + ima	12	8
lcmn + ima	9	8
qua + ima + watn	5	4
pca + ima + watn	10	4
ica + ima + watn	9	5
lcm + ima + watn	11	7
quan + ima + watn	5	2
pcan + ima + watn	11	5
ican + ima + watn	10	6
lcmn + ima + watn	9	6

Based on the result shown in table 7.2 the following feature selection methods were selected for a further inquiry: qua, qua + ima, qua + ima + watn. These three methods were selected to compare between a method using only information from the supressed MRSI, a method that also used image information, and a method that included all three types of data. The methods that used water normalized values were not considered, since water normalization generally did not improve the results. The three selected methods performed best in their class.

The distribution plots for the first two models are shown in figure 7.7 to illustrate the difference of class isolation in feature space between a data set which does not use image data and a data set which does. When image information is used (figure 7.7b), the classes are clearly more isolated. This should improve the classification result of especially CSF, grade II, grade III and meningioma.

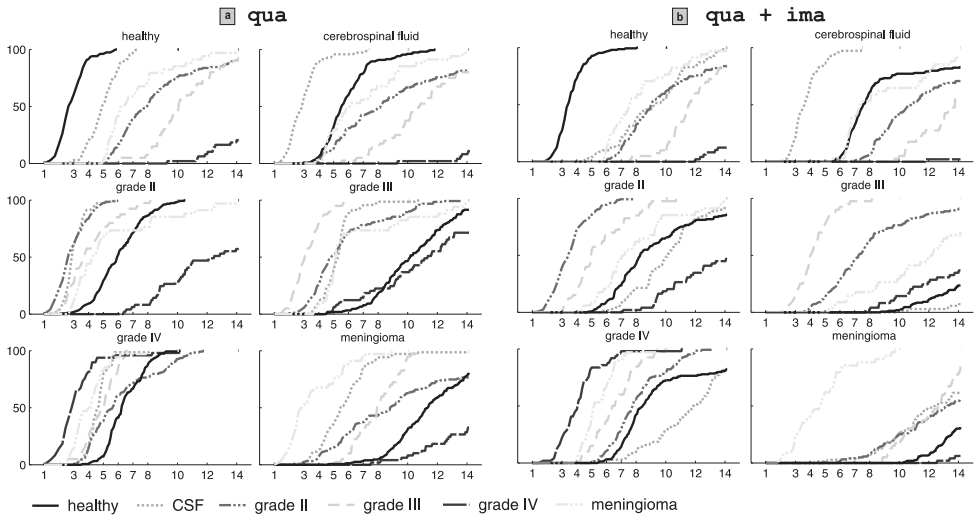


Figure 7.7. Distribution plots of input data from two feature selection methods.

Distribution plots provide information of the overlap of classes in feature space. Each subplot shows the cumulative sum of objects for each class within a Mahalanobis distance to the centroid of one specific class. As example, the subplot with the healthy centroid as origin is explained (figure 7.7a, upper left). When the mahalanobis distance (x-axis) increases, only objects in the healthy class accumulate. This is normal, since the centroid of the healthy class is surrounded by healthy objects. After 3 times the Mahalanobis distance, also

CSF objects are encountered (dotted-gray), followed by meningioma objects. As a result, the healthy class appears to be quite isolated. Other objects in the neighborhood, are mostly from CSF. This implies that sometimes a 'real' CSF object is misclassified as healthy. From the grade II plot in figure 7.7a, it is clear that CSF objects are also in the vicinity of the grade II centroid. Therefore, CSF objects will sometimes be misclassified as grade II. If the same two subplots are viewed in figure 7.7b, then it is immediately clear that objects from other classes are further away from all class centroids. CSF object do not interfere anymore with the healthy of grade II class, and thus will be less misclassified. This is immediately clear when table 7.3 and 7.4 are compared.

To investigate the results of the 3 chosen feature selection methods, their classification tables are shown in tables 7.3, 7.4, and 7.5. When only quantitation is used, the results of CSF, grade III and meningioma are bad. For instance, CSF voxels are misclassified as healthy tissue (1) and grade II (11), which could have been to be expected from the distribution plot depicted in figure 7.6a. Also the misclassification of grade III objects as grade II and meningioma as grade IV is not a surprise. If also image information is used, the classification results increase substantially. CSF is 10 % correctly classified now, as well as classification of the meningiomas. Only misclassifications between grade II and III still occur (one grade II is still misclassified as grade IV). When the water signal is added during feature selection, the classification result slightly improves. Addition of the water signal increases the classification results of all feature selection methods, but only slightly.

Table 7.3. classification result of the test set using the 'qua' method for feature selection.

	healthy	**CSF**	**grade II**	**grade III**	**grade IV**	**MNG**	**α-error**
healthy	65						0
CSF	1	18	11				40
grade II			50	2	1		5.6
grade III			9	8			53
grade IV					21		0
MNG					4	10	28.6
β-error	1.5	0	28.5	20	19.2	0	

Table 7.4. classification result of the test set using the 'qua + ima' method for feature selection.

	healthy	CSF	grade II	grade III	grade IV	MNG	α-error
healthy	65						0
CSF		30					0
grade II			51	1	1		3.8
grade III			6	11			35.3
grade IV					21		0
MNG						14	0
β-error	0	0	10.5	8.3	4.5	0	

Table 7.5. classification result of the test set using the 'qua + ima + watn' method for feature selection.

	healthy	CSF	grade II	grade III	grade IV	MNG	α-error
healthy	65						0
CSF		30					0
grade II			51	1	1		3.8
grade III			5	12			29.4
grade IV					21		0
MNG						14	0
β-error	0	0	8.9	7.7	4.5	0	

Last, the data set using 'qua + ima + watn' as feature selection was classified using selection rules in the classifier. The critical distances for assignment to 'unknown' were estimated from the distribution plot setup by the 'qua + ima + watn' data set. The following values were used: healthy, 6.0; CSF, 6.0; grade II, 6.0; grade III, 6.0; grad IV, 4.5; MNG, 8.0. If an object was assigned to grade II, but the distance to grade III was smaller than 1.3 time the distance to grade II, then the object was classified as 'undecided'. The same rule applied if the object was assigned to grade III. These rules led to the classification as shown in table 7.6.

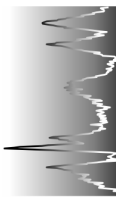


Table 7.6. classification result of the test set using the 'qua + ima + watn' method for feature selection and selection rules in the classifier.

	healthy	CSF	grade II	grade III	grade IV	MNG	unknown	undecided
healthy	61						4	
CSF		28					2	
grade II			50		1		1	1
grade III			3	12				2
grade IV					21			
MNG						14		

Discussion

It is important to know the effect of feature reduction on MRSI data, if this data is going to be used for tumor classification. Feature reduction is obligatory, since the number of objects in each class should be 5 – 10 times higher than the final number of features. It is important to use the available MR information as extensively as possible, since classification is hampered by a low signal to noise ratio of the spectroscopic data, the low number of cases to build the model from and the large number of pathologic different brain tumors. Furthermore, an extensive extraction of class discriminating information can only be performed with a feature extraction method suited for the job. Papers have been published[7, 24] which give a good overview of the preprocessing and feature reduction methods for the classification of MRS data. Finding the best general feature reduction procedure is a difficult issue, because it depends on the particular data used and the specific problem to be solved. Somorjai[25] found that for the classification of thyroid neoplasms principal components analysis on the MR spectra gave similar results as picking peak heights or peak areas. In [26] it was observed that feature reduction with the use of selected frequency regions or peak quantitation resulted in worse classification results for brain tumors as compared to complete spectra. However, they used a support vector machine during classification which deals with the dimensionality problem. Furthermore, their results were based on long echo time (TE = 135 ms) spectra, while in this paper short echo time spectra are used. Long echo time spectra are less noisy, but also contain less

information. Feature reduction could be especially important in the selection of features from short echo time MRS data, because of the ability of most methods to separate information from noise.

We have found that the feature reduction on the MRSI data did not have much effect on the classification result. The quantitation method produced the best result throughout all the investigations. The LCModel approach performed worst. This was not expected, since in theory this model would quantitate the important metabolites better than our 'simple' quantitation approach. The principal component analysis was found to be slightly worse than the quantitation method. This method can be very interesting, since it is easily automated, user independent, and it doesn't need a priori information. Independent component analysis did not perform better than PCA and is more user dependent. Therefore it is not considered a better choice than PCA.

Inter patient normalization on the metabolite MRSI data before feature reduction did not improve classification. However, inter patient normalization as well as intra normalization are an important issue. As shown in figure 7.4, the water image and metabolite images show non-uniformities in the excitation profile, especially at the borders. The lower signal at the borders is probably caused by the profile of the selective pulses of STEAM and the processing of data. These non-uniformities can be partially corrected[27] by means of normalization, which should in principle improve classification. However, these intra-normalization requires a phantom-acquisition with the same spatial dimensions as the metabolite MRSI and was not performed. Inter patient normalization is very important if data from different sites, protocols and MR systems are used[28]. In our case, all data was acquired using the same protocol, site, operator and MR system, and therefore we were not surprised that normalization had no dramatic effect. We could however see improvements in the distribution of the water signals of the voxels in the different classes in the training set, as shown in figure 7.6. Therefore, the normalized water signals were used during combination of data sources.

The classification results dramatically improved if the metabolite MRSI data were combined with the MR image information. To our knowledge this has not been described before in literature. Szabo de Edelenyi[6] uses information from the T_2-weighted image during classification, but did not compare the classification result with classification without MRI information. A comparison of figure 7.7a and b provides a good overview of the classes in multivariate space before and after introduction of the MRI information. All

classes are more isolated in multivariate space if MR image information is used. The CSF class, which was close to the grade II class is located far from all other classes if image information is added. This totally eliminates the α-error (compare table 7.3 and 7.4) of CSF and therefore lowers the chance that a CSF voxel is wrongly assigned to the grade II (or other) class. The meningioma class is close to the grade IV class using only metabolite data (figure 7.7a, grade IV subplot). If also image data is used, the separation is better, and the classification result of meningioma increases dramatically (table 7.4). An overlap between the grade II and III classes still occurs after addition of image information, but the results on the independent test set are better. Since the two classes originate from the same type of (glial) tumor, an overlap is always to be expected.

We show that the classification result of brain tumors greatly improves if the combined power of MRSI and MRI data is used. The addition of image information increases the classification result on an independent data set from ~ 78 % to ~ 92 %. Addition of the MRI information in classification procedures seems to be unavoidable, since MR images obtained with different contrasts are always acquired during standard clinical investigations. Furthermore, it raises the question whether the addition of information from other sources of data, like position of the tumor, age and sex of the patient (clinical information) will further increase the classification results. This might be especially important if more tumor classes are considered at the same time. To use more sources of data however, a robust database of validated cases needs to be setup. Already some databases have been developed[16], and they can be a source for the development of new procedures.

Addition of water amplitude information as calculated from the unsuppressed MRSI data slightly improves classification, although no significant change was detected. However the box plots in figure 7.5 clearly show differences in the water signal for the classes. The use of the water signal has already been recognized long ago as a tumor marker[29] and could contribute to a more accurate brain tumor classification. This also implies that normalization of a voxel with a water signal obtained from the same voxel is dangerous. For inter patient normalization it might be better to calculate a water normalization factor from multiple voxels selected from healthy tissue.

Last, we have shown that the introduction of critical distances around the centroids of the classes in the training set can further improve the classification result. If we apply critical distances in our model, only 4 object of

200 are misclassified. Of 10 objects a classification cannot be given. The classification result with use of an independent test set is 93%, and this is approaching the normally accepted 5% classification error in statistics. If critical distances are introduced, the classes within the feature space are defined by hypervolumes. A new object is only assigned to a class if it falls within its hypervolume. If it falls outside, it is considered to be an outlier, and thus 'unknown'. This introduction of a class of 'unknown' objects is important, since there are much more tumor types than most classifiers can deal with. In this way, a new object from a tumor the classifier cannot handle is not forced into an incorrect class.

Acknowledgement

The EU funded INTERPRET project is gratefully acknowledged. (IST-1999-10310; http://carbon.uab.es/INTERPRET/). The authors acknowledge Leentje Vanhamme for providing the water signal modeling software. Marinette van der Graaf, Yvonne van der Meulen and Mark Rijpkema are acknowledged for acquisition of data and discussions.

References

[1] Meyerand ME, Pipas JM, Mamourian A, Tosteson TD, Dunn JF. Classification of biopsy-confirmed brain tumors using single-voxel MR spectroscopy. *American Journal Neuroradiology* 1999; **20**: 117-123.

[2] Möller-Hartmann W, Herminghaus S, Krings T, Marquardt G, Lanfermann H, Pilatus U, Zanella FE. Clinical application of proton magnetic resonance spectroscopy in the diagnosis of intracranial mass lesions. *Neuroradiology* 2002; **44**: 371-381.

[3] Preul MC. Accurate, noninvasive diagnosis of human brain tumors by using proton magnetic resonance spectroscopy. *Nature Medicine* 1996; **2**(3): 323-325.

[4] Govindaraju V, Young K, Maudsley AA. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR in biomedicine* 2000; **13**: 129-153.

[5] Simonetti AW, Melssen WJ, van der Graaf M, Postma GJ, Heerschap A, Buydens LMC. A Chemometric Approach for Brain Tumor Classification

Using Magnetic Resonance Imaging and Spectroscopy. *Analytical Chemistry* 2003; **75**(20): 5352-5361.

[6] Szabo De Edelenyi F, Rubin C, Estève F, Sylvie G, Décorps M, Lefournier V, Le Bas J-F, Rémy C. A new approach for analyzing proton magnetic resonance spectroscopic images of brain tumors:nosologic images. *Nature Medicine* 2000; **6**(11): 1287-1289.

[7] Hagberg G. From magnetic resonance spectroscopy to classification of tumors. A review of pattern recognition methods. *NMR in biomedicine* 1998; **11**: 148-156.

[8] Howells SL, Maxwell RJ, Griffiths JR. Classification of tumour 1H NMR spectra by pattern recognition. *NMR in Biomedicine* 1992; **5**: 59-64.

[9] Preul MC, Caramanos Z, Leblanc R, Villemure JG, Arnold DL. Using pattern analysis of in vivo proton MRSI data to improve the diagnosis and surgical management of patients with brain tumors. *NMR in biomedicine* 1998; **11**: 192-200.

[10] Klose U. In vivo proton spectroscopy in presence of eddy currents. *Magnetic Resonance in Medicine* 1990; **14**: 26-30.

[11] Pijnappel WWF, van den Boogaart A, de Beer R, van Ormondt D. SVD-based quantitation of magnetic resonance signals. *Journal of magnetic resonance* 1992; **97**: 122-134.

[12] Vandeginste BGM, Massert DL, Buydens LMC, De Jong S, Lewi PJ, Smeyers-Verbeke J. *Handbook of Chemometrics and Qualimetrics: Part B*; Elsevier, 1998.

[13] Wehrens R, Simonetti AW, Buydens LMC. Mixture modelling of medical magnetic resonance data. Journal of chemometrics 2002; **16**(6): 274-282.

[14] Reddick WE, Glass JO, Cook EN, David Elkin T, Deaton RJ. Automated Segmentation and Classification of Multispectral Magnetic Resonance Images of Brain Using Artificial Neural Networks. IEEE transactions on medical imaging 1997; **16**(6): 911-918.

[15] Lukas L, Devos A, Suykens JAK, Vanhamme L, Van Huffel S, Tate A, Majos C, Arús C. The use of LS-SVM in the classification of brain tumors based on magnetic resonance spectroscopy signals. in Proc. of the European Symposium Artificial Neural Networks (ESANN'2002), Bruges, Belgium, Apr. 2002, pp. 131-136.

[16] http://carbon.uab.es/INTERPRET/cdap.html

[17] Simonetti AW, Melssen WJ, van der Graaf M, Heerschap A, Buydens LMC. Automated correction of unwanted phase jumps in reference signals which corrupt MRSI spectra after eddy current correction. *Journal of magnetic resonance* 2002; **159**: 151-157.

[18] Campbell ID, Dobson CM, Williams R.J.P., Xavier A.V., Resolution enhancement of protein MR spectra using the difference between a broadened and a normal spectrum, Journal of Magnetic Resonance 11: 172-181, 1973

[19] A. Hyvärinen and E. Oja. Independent Component Analysis: Algorithms and Applications. Neural Networks, 13(4-5):411-430, 2000.

[20] A. Hyvärinen. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. IEEE Transactions on Neural Networks **10**(3):626-634, 1999.

[21] A. Hyvärinen, J. Särelä and R. Vigário. Bumps and Spikes: Artifacts Generated by Independent Component Analysis with Insufficient Sample Size. In Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99), pp. 425-429, Aussois, France, 1999.

[22] Stepen W. Provencher, Estimation of metabolite concentrations from localized *in vivo* proton NMR spectra. *Magn. Reson. Med.* 1993; **30**: 672-679.

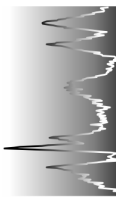
[23] Stephen W. Provencher, Automatic quantitation of localized in vivo 1H spectra with LCModel, NMR Biomed. 2001; 14: 260-264

[24] El-Deredy W. Pattern recognition approaches in biomedical and clinical magnetic resonance spectroscopy: a review. *NMR in Biomedicine* 1997; **10**: 99-124.

[25] Somorjai RL, Nikulin AE, Pizzi N, Jackson D, Scarth G, Dolenko B, Gordon H, Russell P, Lean CL, Delbridge L, Mountford CE, Smith ICP. computerized consensus diagnosis: A classification strategy for the robust analysis of MR spectra. I. Application to ^{1}H spectra of thyroid neoplasms. *Magnetic Resonance in Medicine* 1995; **33**: 257-263.

[26] Lukas L, Devos A, Suykens JAK, Vanhamme L, Howe FA, Majos C, Moreno-Torres A, Van der Graaf M, Tate AR, Arús C, Van Huffel S. Brain Tumour Classification based on long echo proton MRS signals. *International Journal Artificial Intelligence in Medicine* 2003.

[27] Wild JM, Marshall I. Normalisation of metabolite images in ^{1}H NMR spectroscopic imaging. *Magnetic resonance imaging* 1997; **15**(9): 1057-1066.



[28] Tate AR, Majós C, Moreno A, Howe FA, Griffiths JR, Arús C. Automated classification of short echo time in in vivo 1H brain spectra: a multicenter study. *Magnetic Resonance in Medicine* 2003; **49**: 29-36.

[29] Damadian R. Tumor detection by nuclear magnetic resonance. *Science* 1971; **171**: 1151-1153.

Chapter 8

Concluding remarks and recommendations

Arjan W. Simonetti

Chemometric methods have been developed that use the combined power of proton MRSI data and MRI data, to aid in a more reliable and noninvasive brain tumor diagnosis. To obtain a reliable brain tumor classification, all steps in the classification strategy should be tuned and carefully executed. Therefore, it has been investigated what the necessary contributions to the total strategy could be. This has led to the development of chemometric methods which are necessary for the automated preprocessing of the data, for the feature reduction of the data and for the classification procedure. The classification procedure also provides probabilities of classification, which are important in the assessment of the reliability of the classification and might provide information regarding the heterogeneity of the tumor tissue.

The developed chemometric methods have been tested and evaluated within a semi-automated classification system. This system provides an estimate of the location, type and grade of the tumor and can, for example, be used as a clinical tool to locate the best position for a brain tumor biopsy. We have found that the classification of glial tumors of three different grades, meningioma tumors, cerebro-spinal fluid and healthy brain tissue can be performed with an accuracy of up to 93%.

Although feature reduction on the MRSI data is obligatory, the method of feature reduction is not a very crucial factor for successful classification. Four accepted methods of feature reduction (simple quantitation, principal component analysis, independent component analysis and LCModel) have been compared and they don't show dramatic differences. However, the automated preprocessing before the feature reduction is important! Most of the preprocessing procedures used in our models have originally been developed for use in single voxel spectroscopy. If applied automatically on a large set of MRSI spectra this causes problems, which are difficult to overcome. In this thesis a new filter is described to circumvent the occurrence of an artifact that may be introduced in MRSI spectra after Eddy current correction. Furthermore, other processing methods to automatically preprocess MRSI spectra should be developed. For example, the automation of baseline correction and first order phase correction of short echo time MRSI data are highly desirable.

Next to preprocessing, the combination of features from MRI data and MRSI data greatly improves the classification accuracy. This raises the question whether the additional introduction of independently obtained data can further improve the classification result. For example, clinical information

like the age, gender and clinical symptoms can be introduced into the classification model. Radiologists find the approximate location of the brain tumor very important for the diagnosis. This location can be established in an early stage of the investigation and could be a good discriminator.

Also the addition of spectroscopic data from other sources than proton MRSI may improve the classification. The addition of data from other magnetic resonance applications, like ^{31}P-spectroscopy or functional imaging, might be worthwhile to investigate. As mentioned in the introduction, the use of ^{31}P-MRS has been found important in studies of energy metabolism and intracellular pH. Thus, information obtained from ^{31}P spectroscopy can be helpful in the improvement of brain tumor classification. However, ^{31}P-MRS is time consuming and it requires specialized equipment, which is often found only in academic hospitals. Therefore, ^{31}P-MRS in combination with proton MRSI might not be feasible.

The combination of MR data with CT data and PET data is probably more interesting. At the moment, MRI, CT and PET are commonly applied during the investigation of a patient with a brain tumor. All three methods provide an image of the brain, and all with a different information content. Thus, it will be highly desirable to combine these data for brain tumor classification and chemometrics can play a role in the extraction of information of these combined data sources.

In this thesis it has been shown that it is possible to get insight into the reliability of classification during the classification of brain tumors. This is performed by the calculation of a probability of classification for each individual voxel, or for a segment of multiple voxels. This is very important, since the clinician needs an indication of the reliability of the classification result as calculated by the model. Probabilities of classification in tumor diagnosis are rather new and other methods need to be investigated in the near future. A Bayesian approach could be considered, since this provides a way to incorporate prior knowledge into the system.

Further investigations into the nature of the classification models is also an issue to consider. Presently, the spatial information in the maps and images is not used yet to improve the classification. Each voxel or pixel is treated as an independent source that needs to be classified. However, each voxel has several neighbors that have a high probability of belonging to the same class as the voxel under investigation. This information could be used by a classification model to obtain a more reliable/robust classification.

Also classification models that are of help in the follow up of patients can be developed. One could think of using MR data from several time instances to predict the development of the tumor after the patient has been treated. In the nearby future, time series of functional MRI in combination with chemometrics could be very important in the prediction of changes of brain function in the healthy brain tissue surrounding a brain tumor. At the moment, functional MRI is already used to investigate which parts of the brain have an important function during specific tasks like speech, vision and motoric actions. In this way, it is possible to map the function of the brain that surrounds a tumor. If there is loss of brain function in time, then this could be a reason to change the treatment of the patient. It is even better to predict the loss of brain function and to prevent it by changing the treatment.

Another, completely different subject that needs attention, concerns the implementation of the classification software in a clinical environment. To do so, the development of a graphical user interface is highly desirable. The lay-out of a desired user interface has been investigated by the INTERPRET consortium, but a practical implementation of the processing and classification models and the user interface has not been performed yet. Since most of the (pre)processing is automated, the system can be simply integrated in the clinic. However, some considerations have to be made. The classification software has to be robust, so further testing and removal of bugs are needed. It has to be investigated in what clinical environment the clinicians want to use the software. If it is desired, for example, to use the software real time during the investigation of the patient, then other specifications are needed than if the software is to be used after the investigations. In this context one should think of investigations into calculation times or accessibility of the software to the non-expert user.

In future it is also highly desirable to link the software containing the classification strategy to a database that is independent and that contains a high number of cases sampled from all around the world. This database should be maintained by experts in the field of chemometrics, informatics and clinical diagnostics. Such a database will ensure that training sets with a large number of validated samples for several classes and a high quality are available. Furthermore, it avoids that non-experts change the local database of training sets. The INTERPRET consortium has made a start to setup such a database, but at the moment it does not contain a feature to directly link a processed training set to a classification algorithm.

Last, the best resolution of the classification maps should be investigated. Since MR images are obtained with a far higher spatial resolution than MRSI, rigorous processing is needed to equalize the two. Chemometrics can be of help in the development of better methods. Increasing the resolution to the MRI level preserves all the image information, and could therefore increase the classification result and reliability. It also results in classification maps with a high spatial resolution. However, an increase in resolution may also decrease the robustness and reliability of classification, since the spectroscopic information used for the classification of a voxel is acquired from a large volume around that voxel.

I have written this thesis among other things because I am convinced that the use of chemometrics is very valuable for the diagnosis of patients with a brain tumor. Especially in future, since the number of specialized techniques in the clinic will only grow. The data that are generated by these techniques have to be presented to the clinician in the most informative way. To do so, new sophisticated methods have to be developed, that will be present during the whole trajectory of processing. The last four years I have been investigating such methods. I have created an user interface that offers the possibility to quickly process data. The filter described in chapter three is used during processing. With the interface, I have setup a large dataset which is used by several researchers who are developing methods to optimize tumor diagnosis. Furthermore, I have developed a classification model (chapter five) that provides excellent results. It is very well possible to classify tumors and to get information of the reliability of classification as well.

I have also shown that the combination of MRI and MRSI data improves the classification of different tumor grades, healthy tissue and cerebro-spinal fluid. With this result, I hope to convince other researchers to combine data from different sources.

The results of my work have been integrated into a demo. This demo is made to show the clinician how chemometrics can contribute to tumor diagnosis. I hope that this demo convinces clinicians that automated processing of MR data is essential for determination of the correct diagnosis of brain tumors.

Summary

Brain tumor diagnosis is a difficult challenge for the clinician, since it is rarely possible to decide upon the type and grade of the tumor with absolute certainty from the clinical history and classic radiological investigations. To improve the diagnosis of persons with a brain tumor, new radiological methods have been developed in the last decade. In the case of magnetic resonance, the classic magnetic resonance imaging (MRI) has, amongst others, been extended with functional MRI, diffusion weighted MRI and magnetic resonance spectroscopic imaging (MRSI). In the case of MRSI, a cross-section of the brain is measured, in which each pixel (called voxel) contains a spectrum. This spectrum has a very specific shape, which is defined by the tissue it is measured from. Therefore, MRSI could be very important in the investigation of brain lesions.

However, due to the increased number of methods, an overwhelming amount of data is generated, which is not easily accessible by the clinician. Therefore, it is necessary to develop automated data processing and methods based on pattern recognition (chemometrics), that are able to objectively interpret the data. This would provide the clinician with tools to extract relevant information about the disease of the patient, without him spending to much time. Furthermore, methods that use pattern recognition are capable to obtain relevant information from data that would have been very difficult to recognize by the clinician.

Chemometrics is a discipline that has the goal to obtain information and knowledge from data in the field of chemistry. The methods used in chemometrics are of a mathematical and statistical nature and are executed on a computer. These methods are often optimized for the interpretation of multivariate (multi-dimensional) data. The human eye is very good in the interpretations of one and two dimensional data. Often, this is best done on paper as a graph, figure or image. In the case of three dimensional data, the interpretation already becomes more difficult. If even more dimensions are available, then the human mind needs help from chemometrics to interpret

the data. Often, it is tried to extract the best information from the multivariate data and display it in a two dimensional image or graph.

In this thesis, several chemometric methods are investigated, that could contribute to the automated interpretation of large data sets obtained with MRI and MRSI. The goal of the development of these methods is to extract as much relevant information about the type and grade of brain tumors. Eventually, these methods can be used in the construction of a decision support system for brain tumor diagnosis. This program could be used by clinicians to come to a more precise and less time consuming brain tumor diagnosis. This is very important, since the treatment of a brain tumor is very dependent on its type and grade.

The development of chemometric methods that can be used to improve the diagnosis of patients with a brain tumor encompasses several challenges. The data preprocessing needs to be automated, quick, robust, and error propagation needs to be kept to a minimum. Classification models have to be developed to automatically locate and classify the tumor type of a brain tumor. The classification model should also provide an estimate of the reliability of the classification. This reliability of classification is especially important, since the clinicians who will use the system have to be convinced that the classification results are meaningful. Furthermore, the classification result should be spatially resolved and presented image-like, to facilitate an easy interpretation by clinicians in the clinic.

The introduction of this thesis provides the reader with some background on tumor diagnosis. The tumor types investigated in this research are specified and the role of magnetic resonance imaging and spectroscopy in tumor diagnosis is briefly explained. This is followed by a short definition of chemometrics and the role it could play in the classification of brain tumors using MRI and MRS.

In Chapter two an overview is given of the challenges for chemometrics in the analysis of MR images and spectra of brain tumors. The basic aspects of MRI and MRSI are outlined, as well as the processing of images and spectra. The main contributions of chemometrics to the analysis of MR data used for brain tumor classification are outlined. This involves the segmentation of MR images, the processing of spectroscopic data to compensate for instrumental imperfections and the development of classification models.

One of the steps in the preprocessing of MR spectra is called 'Eddy current correction'. Eddy current correction normally improves the interpretation of the spectra, but in the case of MRSI spectra sometimes an artifact is introduced. This artifact can be circumvented by the application of a filter, which is described in Chapter three.

After preprocessing of MRI data and MRSI data, similar voxels of a patient can be clustered to segment the brain in several regions, that have comparable metabolic and morphologic patterns. The development of such a semi-automatic unsupervised clustering technique is described in Chapter four.

The segments found during clustering can be classified using a new classification algorithm. This algorithm also calculates the probability of classification for each segment. Furthermore, the model can also calculate these values for each separate voxel. The results are presented image-wise and can be superimposed on high resolution MR images which provide relevant morphologic information. This is described in Chapter five.

By visual comparison of the constructed probability of classification images, the clinician gets insight in the tumor morphology and type. The two classification images, one from the classified segments and one from the classified individual voxels, provide a consensus diagnosis of the tumor. Chapter six describes a new scientific decision support tool that uses both classification results and the probabilities of classification to aid the clinician in the diagnosis of a patient with a brain tumor.

In Chapter seven, several feature reduction methods are compared in their ability to classify brain tumors of different type and grade, normal tissue and cerebro-spinal fluid. Also the combination of MRI data, MRSI data and information from the water signal is investigated. Although feature reduction is important, especially the combination of MRI and MRSI information is found to be important.

I have written this thesis among other things because I am convinced that the use of chemometrics is necessary for the diagnosis of patients with a brain tumor. Especially in future, since the number of specialized techniques in the clinic will only grow. The data that is generated by these techniques has to be presented to the clinician in the most informative way. To do so, new sophisticated methods have to be developed, that will be present during the whole trajectory of processing. The last four years I have been investigating such methods. I have created an user interface that offers the possibility to construct a totally processed dataset. The filter described in chapter three is

among other things used in this user interface. Using the interface I have collected a large dataset, that is used by several researchers who are developing methods to optimize tumor diagnosis. Furthermore, I have developed a classification model (chapter five) that provides excellent results. It is very well possible to classify tumors and to get information of the reliability of classification as well. Next, I have investigated the influence of several feature selection methods on the classification result of different tumors, healthy tissue and cerebro-spinal fluid. About this topic, not much was yet published in the academic society. With the result, I hope to convince other researchers to combine data from different sources. Moreover, I hope that I have contributed to the awareness of clinicians that automated processing of MR data is very important for determination of the correct diagnosis in the clinic.

Samenvatting

De diagnose van hersentumoren is een moeilijke taak voor de radioloog, omdat het vaak niet mogelijk is om met zekerheid het tumor-type en de graad vast te stellen op basis van de huidige diagnostiek. Om de diagnose van personen met een hersentumor te verbeteren, zijn er in de afgelopen tien jaar verscheidene nieuwe radiologische onderzoekstechnieken ontwikkeld. In het geval van magnetische resonantie imaging (MRI) is de klassieke MRI opname o.a. uitgebreid met functionele MRI, diffusie gewogen MRI en magnetische resonantie spectroscopie (MRSI). In het geval van MRSI wordt er een hersenscan uitgevoerd, waarin zich in elk pixel (voxel genoemd) een heel spectrum bevindt. Dit spectrum heeft een heel specifieke vorm, die wordt bepaald door het weefsel waaruit het is opgenomen. Daardoor kan MRSI dan ook zeer geschikt zijn om het type en de graad van hersentumoren te onderzoeken.

Door het toenemen van het aantal onderzoekstechnieken ontstaat er echter een bijzonder grote hoeveelheid data, die niet altijd door de arts geïnterpreteerd kan worden. Daarom is het noodzakelijk om geautomatiseerde technieken te ontwikkelen die in staat zijn de data objectief te interpreteren. Deze methoden zijn gebaseerd op patroonherkenning (chemometrie). De automatische interpretatie van de grote hoeveelheid data stelt de arts in staat om de meest relevante informatie betreffende de ziekte van de patiënt te overzien, zonder dat hij daar veel tijd voor nodig heeft. Daarnaast zijn methoden die gebruik maken van patroonherkenning in staat om relevante informatie uit de data te verkrijgen die door de arts zeer moeilijk te herkennen zou zijn.

Chemometrie is de discipline die als doel heeft zoveel mogelijk informatie en kennis te vergaren uit data die is opgenomen op het gebied van de chemie. De methoden die door de chemometrie gebruikt worden zijn wiskundig en statistisch van aard en worden uitgevoerd door een computer. Vaak zijn de chemometrische methoden geoptimaliseerd om te werken met multivariate (multivariabele) data. Het menselijk oog is prima in staat om data te interpreteren die uit een of twee variabelen bestaat. Vaak gaat het dan om

grafieken, figuren of plaatjes. In het geval van drie variabelen wordt de interpretatie al moeilijker. Onze wereld is driedimensionaal, maar het interpreteren van een driedimensionaal figuur op papier is al bijna onmogelijk. Als het om nog meer variabelen gaat, dan heeft de mens hulp nodig van onder andere de chemometrie om de data te interpreteren. Vaak wordt er geprobeerd om de multivariate data zodanig te bewerken dat het te projecteren is in een tweedimensionaal plaatje of figuur.

In dit proefschrift worden verschillende chemometrische methoden beschreven die kunnen bijdragen aan de automatische interpretatie van grote datasets die zijn ontstaan door MRI en MRSI metingen. Het doel van de ontwikkeling van deze methoden is het verkrijgen van zoveel mogelijk relevante informatie over het type en de graad van hersentumoren. Uiteindelijk kunnen de ontwikkelde methoden worden gebruikt in de constructie van een beslissingssysteem, dat de arts helpt om in een korte tijd, maar met hoge nauwkeurigheid, de diagnose van een hersentumor vast te stellen. Dit is bijzonder belangrijk omdat de behandeling van patiënten met een hersentumor erg afhankelijk is van het type tumor en de graad van de tumor.

De ontwikkeling van chemometrische methoden, die kunnen worden gebruikt om de diagnose van patiënten met een hersentumor te verbeteren, bevat meerdere uitdagingen. De voorbewerking van de ruwe data moet worden uitgevoerd met snelle, robuuste, geautomatiseerde software, die de opeenstapeling van fouten minimaliseert. Daarnaast moeten er classificatie-modellen worden ontwikkeld die in staat zijn om de plaats van de tumor en het type en de graad te herkennen. Deze modellen moeten ook kunnen aangeven met welke maat van zekerheid de typering heeft plaatsgevonden. Deze maat van zekerheid is zeer belangrijk, omdat de arts overtuigd moet zijn van de kwaliteit van de classificatie. Verder moeten er methoden ontwikkeld worden die de gegevens als een plaatje aanleveren, zodat de resultaten eenvoudig door een arts te interpreteren zijn.

De introductie van het proefschrift geeft de lezer een kort overzicht van tumordiagnose. De tumortypen die gebruikt worden in dit onderzoek worden gespecificeerd en de rol van magnetische resonantie imaging en spectroscopie in de tumordiagnose worden verklaard. Vervolgens wordt een samenvatting gegeven van de definitie van chemometrie en de rol die deze kan spelen in de classificatie van hersentumoren.

In hoofdstuk twee wordt ingegaan op de uitdagingen voor de chemometrie in de analyse van magnetische resonantie imaging en spectroscopie van hersentumoren. De basisaspecten van MRI en MRSI worden omschreven, net als een aantal voorbewerkingtechnieken. Voorts wordt de huidige bijdrage van de chemometrie in de analyse van MR data behandeld. Het betreft hier de segmentatie van MR images, de compensatie van artefacten geïnduceerd door de meetmethode en de ontwikkeling van classificatie modellen.

Eén van de stappen in de voorbewerking van MR spectra wordt 'Eddy current correctie' genoemd. Deze methode verbetert normaal gesproken de interpretatie van spectra, maar bij het gebruik in MRSI kan er soms een artefact geïnduceerd worden. Dit artefact kan worden geëlimineerd door het gebruik van een filter. Dit filter wordt beschreven in hoofdstuk drie.

Na de voorbewerking van de MRI en MRSI data, kan deze gebruikt worden om een gesegmenteerd beeld van de hersenen te verkrijgen. De delen van de hersenen die gelijke metabole eigenschappen hebben, worden samengevoegd. De ontwikkeling van een segmentatie algoritme dat gestuurd wordt door een clusteranalyse, wordt beschreven in hoofdstuk vier. Het is een belangrijke stap in de condensatie van de data en het objectief indelen van de hersenen in regio's met dezelfde metabole kenmerken.

De segmenten die tijdens de clusteranalyse gevonden worden, kunnen worden geclassificeerd met gebruik van een nieuw ontwikkeld programma. Dit algoritme berekent ook de statistische betrouwbaarheid van classificatie voor elk segment. Daarnaast kan het model deze waarden ook voor elk los voxel in de MRSI plaatjes berekenen. De resultaten worden als classificatieplaatjes gepresenteerd en kunnen worden geprojecteerd op MR images die een hogere resolutie hebben. Dit wordt beschreven in hoofdstuk vijf.

De twee classificatieplaatjes kunnen tezamen met de betrouwbaarheid van de classificatie gebruikt worden als een consensus diagnose. Wanneer een gebied in de hersenen in beide plaatjes hetzelfde wordt geclassificeerd, dan is de kans groot dat deze classificatie correct is. Als daarnaast ook de statistische betrouwbaarheid hoog is, dan geeft dit nog meer vertrouwen in een correcte weergave van de werkelijkheid. In hoofdstuk zes wordt een beslissingssysteem beschreven dat gebruik maakt van bovenstaande modellen. Dit systeem kan in de kliniek gebruikt worden om de arts te helpen bij de diagnose van patiënten met een hersentumor.

In hoofdstuk zeven wordt een aantal technische methoden met elkaar vergeleken die gebruikt kunnen worden om de meest relevante informatie uit een grote dataset te isoleren. De methoden worden getest in hun vermogen om drie verschillende typen hersentumoren, normaal hersenweefsel en hersenvloeistof te onderscheiden. Daarnaast wordt onderzocht of de combinatie van data, verkregen uit MRI én MRSI, leidt tot een verbetering van de classificatie.

Deze thesis is mede tot stand gekomen omdat ik ervan overtuigd ben dat chemometrische methoden noodzakelijk zijn bij de diagnose van patiënten met een hersentumor. Zeker in de toekomst, aangezien het aantal technisch hoogstaande meettechnieken in de kliniek alleen maar zal toenemen. De data die door deze meetmethoden wordt gegenereerd, moet zo informatief mogelijk aan de arts worden aangeboden. Daarvoor zijn geavanceerde dataverwerkingstechnieken nodig. Deze technieken moeten aanwezig zijn in het gehele traject van de dataverwerking. De laatste vier jaar heb ik binnen dit traject onderzoek gedaan. Ik heb een user interface ontwikkeld die het mogelijk maakt om gemakkelijk een geheel verwerkte dataset te construeren. Daarvoor heb ik onder andere een filter ontwikkeld dat beschreven wordt in hoofdstuk drie. Een grote dataset die door mij is opgebouwd, wordt intussen gebruikt bij een aantal onderzoeken die als doel hebben tumordiagnose te optimaliseren. Daarnaast heb ik een classificatie model ontwikkeld (hoofdstuk vijf) dat tot goede resultaten leidt. Het is zeer wel mogelijk om tumoren te classificeren en daarnaast een beeld te krijgen van de betrouwbaarheid van classificatie. Daarnaast heb ik laten zien wat de invloed is van verschillende dataverwerkingstechnieken op het classificatieresultaat van verschillende groepen tumoren, gezond weefsel en hersenvloeistof. Hierover was in de academische wereld nog niet veel gepubliceerd. Met dit resultaat hoop ik andere wetenschappers te overtuigen om verschillende bronnen van data te combineren. Daarnaast hoop ik met mijn resultaten, die in de vorm van een demo toegankelijk zijn (zie hoofdstuk zes voor een voorbeeld), bij te dragen aan de bewustwording van de arts dat automatische MR informatieverwerking bijzonder belangrijk is bij het stellen van de juiste diagnose in de kliniek.

Abbreviations

^{1}H-MRSI	proton magnetic resonance spectroscopic imaging
2D	two-dimensional
3D	three-dimensional
AIC	akaike's information criterion
ATP	adenosine triphosphate
BIC	bayesian information criterion
Cho	choline-containing compounds
Cr	phosphocreatine plus creatine
CSF	cerebro-spinal fluid
CT	computed tomography
ECC	Eddy current correction
EM	expectation maximization
FID	free induction decay
FWHM	full width at half maximum
Gd-DTPA	Gadolinium-DTPA
GD-image	Gd-DTPA enhanced image
Glc	glucose
Gln	glutamine
Glu	glutamate
HLSVD	hankel lanczos singular value decomposition
ICA	independent component analysis
Ins	myo-inositol
Lac	lactate
LDA	lineair discriminant analysis
Lip	lipids

MNG	meningioma
MR	magnetic resonance
MRI	magnetic resonance imaging
MRS	magnetic resonance spectroscopy
MRSI	magnetic resonance spectroscopic imaging
NAA	N-acetylaspartate
PCA	principal component analysis
PC	principal component
PD-image	proton density weighted image
PET	positron emission tomography
rCBV	regional cerebral blood volume
SVS	single voxel spectroscopy
STEAM	stimulated echo acquisition mode
TE	echo time
TM	mixing time
TR	repetition time
WHO	world health organization

Dankwoord

Beste lezer. U kijkt waarschijnlijk in het dankwoord van dit proefschrift, omdat u graag wilt weten of uw naam ook wordt genoemd. Mocht dit niet zo zijn, wees dan gerust. Ik dank u bij deze alleen al omdat u de moeite heeft genomen mijn proefschrift door te bladeren. Over het algemeen is een proefschrift zo specialistisch dat de meeste mensen niet de vakkundigheid hebben om zich te verdiepen in de inhoud. Speciaal voor diegenen heb ik er alles aan gedaan om de samenvatting en introductie zo begrijpelijk mogelijk te houden. Probeert u ze eens te lezen!

De hoofdstukken waarin het onderzoek van de laatste vier jaar wordt beschreven had ik niet tot stand kunnen brengen zonder de hulp van een groot aantal mensen. Deze wil ik dan ook graag persoonlijk noemen.

Allereerst Lutgarde Buydens en Arend Heerschap: jullie wil ik speciaal bedanken voor het mogelijk maken van dit proefschrift. Binnen de structuur van jullie afdelingen heb ik ontzettend veel kunnen leren en is dit proefschrift gegroeid. Lutgarde, je hebt je enorm ingezet om mij academisch te scholen. Die ene keer dat ik onzin uitkraamde op een (belangrijke) vergadering werd mij vergeven. Maar ik moest er wel voor zorgen dat het niet weer gebeurde!

Daarnaast wil ik alle collega's van de afdeling analytische chemie en de onderzoeksgroep van radiologie ontzettend bedanken. Brigitte, als secretaresse onverslaanbaar; mijn 'interpret' maatjes Willem en Geert; Ron die altijd tijd had voor een vraag. Vooral ook Uwe en Jos. Met jullie heb ik een bijzonder leuke tijd gehad. En als je onderzoek doet, is dat belangrijk. Het zit namelijk niet altijd mee! Met jullie als mede AIO's, wist ik zeker dat het goed ging komen. Ook wil ik Marinette bedanken. Zij stond altijd voor mij klaar; zeer kritisch en altijd deskundig. Verder Dennis "tennis', Mark en Yvonne die de spectro deden, Sjaak voor de Unix-les en alle ISMRM (Philidelphia-Denver-Hawaï) gangers.....het was gezellig!

Naast de collega's op de afdelingen in Nijmegen, wil ik ook alle personen bedanken waarmee ik heb samengewerkt in het Europese interpret project. Vooral Chantal, Joshua, Leentje, Sabine, Carles en Christoph wil ik persoonlijk bedanken. Door deze mensen en het project heb ik vele mooie plekken mogen bezoeken, heb ik veel geleerd en heb ik nu een leuke baan in Frankrijk.

Tot slot wil ik mijn familie bedanken. Alhoewel mijn ouders niet altijd begrepen waar het nu precies over ging, toonden ze altijd belangstelling en waren ze bezorgd over de voortgang! Dat was fijn. Bedankt Pa en Ma!

Karijn, we hebben ons samen door het schrijven van het proefschrift heen geslagen. Zeker de laatste maanden, toen de laatste loodjes het spreekwoord eer aandeden, heb ik haar tekort gedaan. Wou ze met mij iets leuks gaan doen, moest ik weer werken! En dat terwijl buiten de bergen en de sneeuw ons toelachten! Nu hebben we weer meer tijd om er samen op uit te trekken. En dat zullen we nog vele jaren blijven doen.

Gegroet en tot wederziens,

Arjan

Publications

B.P.J. van der Sanden, A. Heerschap, L. Hoofd, **A.W. Simonetti**, K. Nicolay, A. van de Toorn, W.N.J.M. Colier, and A.J. van der Kogel, Effect of carbogen breathing on the physiological profile of human glioma xenografts. *Magnetic Resonance in Medicine* 1999; **42**(3): 490-499.

B.P.J. van der Sanden, A. Heerschap, **A.W. Simonetti**, P.F.J Rijken, H.P.W. Peters, G. Stuben, and A.J. van der Kogel, Characterization and validation of noninvasive oxygen tension measurements in human glioma xenografts by ^{19}F-MR relaxometry. *International Journal of Radiation Oncology Biology Physics* 1999; **44**(3): 649-658.

E. Janssen, P.P. Dzeja, F. Oerlemans, **A.W. Simonetti**, A. Heerschap, A. de Haan, P.S. Rush, R.R. Terjung, B. Wieringa, and A. Terzic, Adenylate kinase 1 gene deletion disrupts muscle energetic economy despite metabolic rearrangement. *EMBO Journal* 2000; **19**(23): 6371-6381.

H. Witjes, **A.W. Simonetti**, and L.M.C. Buydens, Better brain imaging with chemometrics. *Analytical chemistry* 2001; **73**(19): 548A-556A.

A.W. Simonetti, W.J. Melssen, M. van der Graaf, A. Heerschap, and L.M.C. Buydens, Automated correction of unwanted phase jumps in reference signals which corrupt MRSI spectra after eddy current correction. *Journal of magnetic resonance* 2002; **159**: 151-157.

R. Wehrens, **A.W. Simonetti**, and L.M.C. Buydens, Mixture modelling of medical magnetic resonance data. *Journal of chemometrics* 2002; **16**(6): 274-282.

A.W. Simonetti, W.J. Melssen, M.v.d. Graaf, G.J. Postma, A. Heerschap, and L.M.C. Buydens, A Chemometric Approach for Brain Tumor Classification Using Magnetic Resonance Imaging and Spectroscopy. *Analytical chemistry*, 2003; **75**(20): 5352 - 5361.